UNIVERSITÉ DE MONTPELLIER

Mémoire
en vue d'obtenir une

HABILITATION À DIRIGER DES RECHERCHES

Modélisation statistique et développement de méthodes de simulation de processus extrêmes pour des applications en sciences de l'environnement.

présentée par

Gwladys Toulemonde

et soutenue le 16 octobre 2020 à l'Université de Montpellier

devant le jury composé de

Jean-Noël Bacro	Professeur, Université de Montpellier	Examinateur
Liliane Bel	Professeure, AgroParisTech	Examinatrice
Anthony Davison	Professeur, EPFL	Rapporteur
Clément Dombry	Professeur, Université de Besançon	Examinateur
Armelle Guillou	Professeure, Université de Strasbourg	Examinatrice
Anne Laurent	Professeure, Université de Montpellier	Examinatrice
Valérie Monbet	Professeure, Université de Rennes	Rapporteure
Philippe Naveau	Directeur de recherche, CNRS	Examinateur

Table des matières

Première partie

Liste des publications et communications

Tous les articles sont disponibles sur la page https://imag.umontpellier.fr/∼toulemonde/. De plus, les 5 articles **[G1, G3, G6, G7, G20]** sont annexés au document en partie III.

- **Articles parus (13)**

[G1] Carreau J., Toulemonde G. Spatial dependence structure for flood-risk rainfall. Spatial Statistics (2020). In Press. DOI : 10.1016/j.spasta.2020.100410.

[G2] Aouni, J., Bacro, J.N., Toulemonde, G., Colin, P., Darchy, L., Sebastien, B. Design Optimization for dose finding trials : A review. Journal of Biopharmaceutical statistics (2020). In Press. DOI : 10.1080/10543406.2020.1730874.

[G3] Bacro, J.N., Gaetan, C, Opitz, T., Toulemonde, G. Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. Journal of the American Statistical Association (2019). In Press. DOI : 10.1080/01621459.2019.1617152.

[G4] Aouni, J., Bacro, J.N., Toulemonde, G., Colin, P., Darchy, L., Sebastien, B. Assessing dunnett and mcp-mod based approaches in two-stage dose-finding trials. Biostatistics and Health Sciences (2019), 1, 72-88. DOI : 10.21494/ISTE.OP.2019.0397.

[G5] Aouni, J., Bacro, J.N., Toulemonde, G., Sebastien, B. Utility-based dose-finding in practice : some empirical contributions and recommendations. Annals of Biostatistics & Biometric Applications (2019), 3(1). DOI : 10.33552/ABBA.2019.03.000552.

[G6] Chailan, R., Toulemonde, G., Bacro, J.N. A semiparametric method to simulate bivariate space-time extremes. Annals of Applied Statistics (2017), 11, 1403-1428. DOI : 10.1214/17-AOAS1031.

[G7] Bacro, J.N., Gaetan, C, Toulemonde, G. A flexible model for spatial extremes. Journal of Statistical planning and inference (2016), 172. DOI : 10.1016/j.jspi.2015.12.002

[G8] Toulemonde, G., Guillou, A., Naveau, P. Particle filtering for Gumbel-distributed daily maxima of methane and nitrous oxyde, Environmetrics (2013), 24, 51-63. DOI : 10.1002/env.2192.

[G9] Bacro, J.N., Toulemonde, G. Measuring and modelling multivariate and spatial dependence of extremes. Journal de la Société Francaise de Statistique, numéro spécial Extrêmes (2013), 154, 139-155.

[G10] Toulemonde, G., Guillou, A., Naveau, P., Vrac, M., Chevallier, F. Autoregressive models for maxima and their applications to CH4 and N2O, Environmetrics (2010), 21, 189-207. DOI : 10.1002/env.992.

[G11] Beirlant, J., Guillou, A., Toulemonde, G. Peaks-Over-Threshold modeling under random censoring, Communications in Statistics -Theory and Methods (2010), 39, 1158-1179. DOI : 10.1080/03610920902859599.

[G12] Falk, M., Guillou, A. Toulemonde, G. A LAN based Neyman smooth test for Pareto distributions, Journal of Statistical Planning and Inference (2008), 138, 2867-2886. DOI : 10.1016/j.jspi.2007.10.007.

[G13] Ryan, J. Carriere, I. Ritchie, K. Stewart, R. Toulemonde, G. Dartigues, J.F. Tzourio, C. Ancelin, M.L. Late-life depression and mortality : influence of gender and antidepressant use, The British Journal of Psychiatry (2008), 192, 12-18. DOI : 10.1192/bjp.bp.107.039164.

- **Chapitres d'ouvrage (2)**

[G14] Toulemonde, G., Carreau, J., Guinot, V., Space-time simulations of extreme rainfall : why and how ? in S. Manou-Abi, S. Dabo-Niang, J. Salone (eds), Mathematical Modeling of Random and Deterministic Phenomena (2020), Wiley.

[G15] Toulemonde, G, Ribereau, P, Naveau, P. Applications of Extreme Value Theory to environmental data analysis, in M. Chavez, JU. Fucugauchi, M. Ghil (eds), AGU monograph on Extreme Events : Observations, Modeling and Economics (2015), Wiley.

- **Proceedings (4)**

[G16] Palacios-Rodriguez F., Toulemonde G., Carreau J., Opitz T. Space-time extreme processes. Simulation for flash floods in Mediterranean France. Proceedings of the 9th Workshop on spatio-temporal modeling (METMA IX), Montpellier, (2018), 100-103.

[G17] Carreau, J., Toulemonde, G. Extra-parametrized extreme-value copula : an extension to a spatial framework. Proceedings of the 9th Workshop on spatio-temporal modeling (METMA IX), Montpellier, (2018), 104-107.

[G18] Chailan, R. Toulemonde, G., Bouchette, F., Laurent, A., Sevault, F., Michaud, H. Spatial assessment of extreme significant waves heights in the Gulf of Lions, ICCE proceedings (2014), 34.

[G19] Chailan, R., Laurent, A. Bouchette, F. Dumontier, C. Hess, O. Lobry, O. Michaud, H., Nicoud, S. Toulemonde, G. High Performance Pre-Computing : Prototype application to a Coastal Floofing Decision Tool, KSE'2012 : 4th International Conference on Knowledge and Systems Engineering, Danang, Viêt Nam (2012), 195-202.

- **Articles soumis ou en révision (4)**

[G20] Palacios-Rodriguez F., Toulemonde G., Carreau J., Opitz T. Generalized Pareto processes for exploring and simulating space-time extremes : appplication to rainfall data. Soumis (2019). hal-02136681v2.

[G21] Aouni, J., Bacro, J.N., Toulemonde, G., Colin, P., Darchy, L. On the use of utility functions for optimizing phase II/phase III seamless trial designs. Soumis (2019). hal-02491531.

[G22] Aouni, J., Bacro, J.N., Toulemonde, G., Colin, P., Darchy, L. Utility-based dose selection for phase II dose-finding studies. Soumis (2019). hal-02491551.

[G23] Sous, D., Bouchette, F., Doerflinger, E., Meulé, S., Certain, R., Toulemonde, G. On the fractal geometrical structure of a living coral reef barrier. En révision à *Earth Surface Processes and Landforms*, (2020).

- **Articles en préparation (5)**

[P-1] Bacro, J.N., Gaetan, C., Opitz, T., Toulemonde, G. Multivariate Pareto models for threshold exceedances based on exponential gamma ratios.

[P-2] Naveau, P., Opitz, T., Toulemonde, G. Hierarchical time modeling of data including extremes with an application to rainfall data.

[P-3] Bacro, J.N., Carreau J., Gaetan, C., Toulemonde G. Non-stationnary hybrid spatial model for extremes.

[P-4] Palacios-Rodriguez F., Opitz, T., Toulemonde G. Exceedance-based risk measures for multivariate extremes.

[P-5] Palacios-Rodriguez F., Bacro J.N., Di Benardino E., Toulemonde G. On risk measures based on a general model for bivariate tail probabilities.

- **Conférences internationales invitées (11)**

[C1] Toulemonde, G. Climate extremes, CMStatistics, London, UK (décembre 2020).

[C2] Toulemonde, G. Space-time modelling and simulation of extreme rainfall, MASCOT NUM 2020 meeting, Aussois (avril 2021 - initialement prévu mai 2020).

[C3] Toulemonde, G., Bacro, J.N., Gaetan, C., Naveau, P., Opitz, T. A space-time process for extremes : application to precipitation data, ERCIM-CMStatistics, London, UK (décembre 2019).

[C4] Toulemonde, G., Bacro, J.N., Gaetan, C., Naveau, P., Opitz, T. Hierarchical time modeling of data including extremes, Extreme Value Analysis Conference, Zagreb, Croatie (juillet 2019).

[C5] Toulemonde, G., Bacro, J.N., Gaetan, C., Opitz, T. Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data, CMStatistics, Pise, Italie (décembre 2018).

[C6] Toulemonde, G. Sur la modélisation et la simulation de champs spatio-temporels extrêmes, CIMOM'18, Mayotte (novembre 2018).

[C7] Toulemonde, G., Bacro, J.N., Gaetan C. Spatial dependence issues for spatial extremes. Workshop on Extremes, Copulas and Actuarial Science, CIRM, Marseille, France (2016).

[C8] Toulemonde, G., Bacro, J.N., Gaetan C. A flexible dependence model for spatial extremes. 60th World Statistics Congress, International Statistical Institute, Rio de Janeiro, Brésil (2015).

[C9] Toulemonde, G., Guillou, A., Naveau, P. Hidden Markov models for Gumbel maxima. Young researchers school on analyzing extreme events distributions, Aussois (2010).

[C10] Toulemonde, G., Guillou, A., Naveau, P. State-space models in extreme value theory. The 20th Annual Conference of The International Environmetrics Society (TIES), Bologna, Italie (2009).

[C11] Toulemonde, G., Guillou, A., Naveau, P., Vrac, M., Chevallier, F. Auto-Regressive models for maxima with applications to atmospheric chemistry. Statistical Modeling of Extremes in Data Assimilation and Filtering Approaches, Strasbourg (2008).

- **Conférences nationales invitées (3)**

[C12] Toulemonde, G., Bacro, J.N. Modèles asymptotiquement indépendants pour les extrêmes spatiaux. 7èmes Rencontres Statistiques de Rochebrune (2012).

[C13] Toulemonde, G., Guillou, A., Naveau, P. Modèles de Markov cachés pour des maxima : enjeux et applications en sciences de l'atmosphère. 6èmes Rencontres Statistiques de Rochebrune (2010).

[C14] Toulemonde, G., Guillou, A., Naveau, P., Vrac, M., Chevallier, F. An Auto-Regressive model for maxima : Application to atmospheric chemistry. Journées MAS de la SMAI, Rennes (2008).

- **Conférences internationales avec comité de lecture (3)**

[C15] Toulemonde, G., Bacro, J.N., Gaetan, C., Opitz, T. Modelling spatio-temporal extremal dependencies for hourly precipitations in Southern France, ISI, Kuala Lumpur, Malaysia (2019).

[C16] Toulemonde, G., Bacro, J.N., Gaetan, C. Dependence structures for spatial extremes. The 11th International conference on Operations Research (ICOR), Havana, Cuba (2014).

[C17] Toulemonde, G., Guillou, A., Naveau, P. Hidden Markov models for Gumbel maxima. The 21th Annual Conference of The International Environmetrics Society (TIES), Isla de Margarita, Venezuela (2010).

Deuxième partie

Analyse de travaux scientifiques choisis et projet de recherche

Chapitre 1

Introduction

L'axe central de mes activités de recherche en **statistique** est la **théorie des valeurs extrêmes**. J'ai d'abord effectué ma thèse de doctorat, soutenue en 2008, à l'Université Paris 6 dans ce domaine en me concentrant sur des problématiques univariées. Dans un premier temps, dans **[G12]**, j'ai proposé un test lisse d'ajustement à la famille de Pareto motivé par la théorie de Le-Cam sur la normalité asymptotique locale (LAN). J'en ai établi le comportement asymptotique sous l'hypothèse que l'échantillon provient d'une distribution de Pareto et sous des alternatives locales, me plaçant ainsi dans le cadre LAN. Dans une autre partie, j'ai proposé un estimateur des paramètres de la distribution de Pareto généralisée dans le cadre de données censurées aléatoirement à droite. J'ai établi la normalité asymptotique de cet estimateur et j'ai illustré, sur simulations, son comportement à distance finie et l'ai comparé à celui de l'estimateur du maximum de vraisemblance (voir **[G11]**). Enfin la dernière contribution de ma thèse **[G10]** est la proposition d'un modèle linéaire autorégressif adapté à la loi de Gumbel pour prendre en compte la dépendance dans les maxima. J'ai établi des propriétés théoriques de ce modèle et j'ai, par simulations, illustré son comportement à distance finie. Enfin, comme des applications concrètes en sciences de l'atmosphère motivaient ce modèle, je l'ai utilisé pour modéliser des maxima de dioxyde de carbone et de méthane.

Lors de mon recrutement à Montpellier en 2009 en qualité de maitre de conférences, je rejoignais une équipe dans laquelle la thématique des valeurs extrêmes était déjà présente et je bénéficiais alors d'un environnement très stimulant. J'ai pu ainsi dès l'année de mon recrutement participer à la rédaction de deux projets de recherche : un projet ANR et un projet de l'appel à projet "Gestion et impacts du changement climatique" du ministère de l'écologie et du développement durable (MEDD). Ce fut alors l'occasion **d'étendre mes thématiques de recherche et d'orienter mes axes d'études aux cas multivarié, temporel et spatial.** Les concepts clés sont alors totalement différents du cas univarié étudié pendant ma thèse, et il s'agissait plus d'une reconversion thématique que d'un simple élargissement de mon champ d'étude. Ces deux projets m'ont également permis d'inscrire plus solidement ma recherche en sciences de l'environnement et de m'intéresser en particulier à des problématiques climatiques. Pour montrer l'apport des valeurs extrêmes en sciences de l'environnement de manière générale, j'ai également eu l'occasion de participer à un ouvrage collectif par la rédaction d'un chapitre **[G15]**.

Dans un cadre à la fois temporel et multivarié, j'ai mené un travail de **reconstruction de séries temporelles** de maxima de deux gaz à effet de serre, le méthane et l'oxyde nitreux. Les modèles à espace d'états usuels ne sont pas adaptés pour reproduire des dynamiques de données extrêmes. Nous appuyant sur **[G10]**, nous avons proposé dans **[G8]** un tel modèle, ayant l'avantage, outre d'être parfaitement adapté à des maxima, de rester linéaire et d'être simple d'interprétation. La reconstruction de la série cachée a donné de très bons résultats grâce notamment à l'utilisation de poids optimaux dans un algorithme de filtrage particulaire.

Ce travail sur la proposition d'un modèle à espaces d'états pour les extrêmes **[G8]** a également suscité chez moi un **intérêt prononcé pour la notion d'indépendance asymptotique**. On parle d'indépendance asymptotique quand la force de dépendance extrémale diminue à des niveaux élevés pour disparaître finalement par opposition à la dépendance asymptotique qui correspond à une situation de stabilité de la dépendance extrémale quel que soit le niveau extrême considéré. Ces notions sont rigoureusement définies dans le chapitre 2 dédié à la prise en compte de l'indépendance asymptotique dans des modèles spatiaux ou spatio-temporels. Après l'écriture d'une revue **[G9]** sur le thème de la dépendance extrémale du cas multivarié au cas spatial, j'ai co-encadré avec J.N. Bacro la thèse de Néjib Dalhoumi soutenue en 2017 portant sur la modélisation multivariée des queues de distributions suffisamment flexible pour s'adapter à des situations de dépendance et d'indépendance asymptotique.

J'ai poursuivi cette recherche de prise en compte du cas de l'indépendance asymptotique dans des modèles flexibles ou dédiés dans un cadre spatial et spatio-temporel **[G3, G7]** (voir chapitre 2) notamment parce que les précipitations présentent cette caractéristique (Davison *et al.* , 2013; Thibaud *et al.* , 2013; Le *et al.* , 2018). La négliger à tort revient alors à faire une erreur de modélisation de la structure de dépendance et conduit à des biais dans l'évaluation de risques associés. Une autre difficulté dans la modélisation de la dépendance extrémale spatiale est la prise en compte de la **non-stationnarité** dans la structure de dépendance spatiale. Pour répondre à ce besoin, j'ai proposé dans des travaux avec Julie Carreau **[G1, G17]** une modélisation basée sur un mélange de copules de Gumbel spatialisé avec une application sur des données de pluie dans la région des Cévennes.

Outre une modélisation de la dimension spatiale des données, prendre en compte simultanément la **dépendance temporelle** est un véritable défi. Actuellement très peu de modèles le permettent (voir par exemple Davis *et al.* , 2013a,b; Huser & Davison, 2014) et restent, pour la plupart, très difficile à interpréter. Nous avons cherché dans **[G3, G6, G16, G20]** à prendre en compte cette dimension temporelle en plus de la dimension spatiale. Dans ce contexte, travailler sur les dépassements plutôt que sur les maxima rend l'interprétation plus aisée et naturelle. Ces questions d'interprétation deviennent essentielles dès lors que notre objectif principal est un **objectif de simulation d'événements extrêmes.**

En effet, il peut être très intéressant de générer des champs de précipitations ou de vagues pour lesquels on s'attend par exemple à observer une valeur donnée en un site une fois tous les 100 ans. Ces simulations peuvent servir à alimenter des modèles d'impact pour étudier l'effet de ces "scénarios catastrophes". C'est dans ce cadre que j'ai, en 2011, initié une collaboration avec le Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), Géosciences Montpellier et IBM Montpellier. Cette collaboration m'a conduit à co-encadrer la thèse Cifre de Romain Chailan (2012-2015) et à m'intéresser de plus près à l'application du calcul

scientifique et de l'analyse statistique à la gestion du risque inondation en milieu littoral (voir aussi **[G6, G18, G19]**).

C'est en co-encadrant la thèse de R. Chailan au cours de laquelle nous avons proposé une approche semi-paramétrique pour la simulation de processus spatio-temporels de vagues que j'ai réalisé le déficit de méthodes permettant des **simulations réalistes d'extrêmes** c'est-à-dire dont la simulation peut réellement correspondre à un événement observable. C'est pourquoi à partir de 2015 environ, j'ai cherché à mieux comprendre ce manque, fait de cette problématique une de mes orientations de recherche privilégiée (voir chapitre 3) en m'y impliquant fortement et en proposant notamment des projets de recherche structurants sur ce sujet.

Le plus structurant d'entre eux est le projet appelé **CERISE** financé sur 3 années (2016-2018) par l'action MANU (méthodes mathématiques et numériques) du programme LEFE (Les Enveloppes Fluides et l'Environnement) dont l'Institut National des Sciences de l'Univers (INSU) relevant du CNRS est un des principaux financeurs. Ce projet intitulé *Simulation de scenarii intégrant des champs extrêmes spatio-temporels avec éventuelle indépendance asymptotique pour des études d'impact en science de l'environnement* engageait 7 jeunes chercheurs (3 doctorants et 4 jeunes chercheurs recrutés entre 2009 et 2014) sur 11 participants regroupés en 6 partenaires que sont l'Université de Montpellier, l'INRAE (Avignon), l'IRD (Montpellier), l'Université d'Avignon, l'Université de Lyon 1 et l'Université de Venise en Italie. Le principal objectif de CERISE était de simuler des processus spatio-temporels pour les événements extrêmes qui reproduisent la variabilité spatiale et temporelle des processus environnementaux. Plusieurs verrous avaient alors été identifiés parmi lesquels **i) la présence d'indépendance asymptotique ; ii) la nécessité de prise en compte de la dépendance temporelle ; iii) la nécessité d'une interprétation évènementielle des scénarios simulés**. L'objectif n'était pas de proposer un unique modèle levant tous les verrous simultanément mais de nous efforcer dans nos constructions et généralisations de modèles de garder à l'esprit ces questions scientifiques soulevées par l'analyse même des données et par l'expérience des praticiens. En fin d'Introduction, je préciserai de quelle façon les travaux que j'ai choisi de présenter dans le cœur de ce document lèvent ces verrous.

Dans CERISE, les développements ont porté principalement sur l'étude de processus spatiaux et spatio-temporels adaptés aux événements extrêmes permettant pour certains l'indépendance asymptotique. CERISE a permis de mettre en lumière la grande difficulté à générer des champs extrêmes et a renforcé le besoin déjà identifié d'être en capacité d'intégrer des simulations d'événements extrêmes dans des simulations à plus longue durée ou sur des zones spatiales plus importantes. Dans la continuité de CERISE, j'ai obtenu le financement d'un autre projet de recherche LEFE-MANU intitulé *FoRçAges de précipitations par simulation stochastique pour études d'Impacts hydrologiques : des périodes Sèches aux événements Extrêmes* (**FRAISE**) sur la période 2019-2021. Le consortium est plus large que dans *CERISE* et regroupe 14 chercheurs issus de 9 partenaires que sont l'AgroParisTech, le CNRS, l'INRAE (Avignon), l'Inria, l'IRD (Montpellier), l'Université d'Avignon, l'Université de Lyon 1, l'Université de Montpellier et l'Université de Venise en Italie. Riche des enseignements de CERISE et des récentes avancées en modélisation spatio-temporelle des valeurs extrêmes, nous souhaitons proposer de nouvelles constructions de générateurs stochastiques spatiaux incluant des champs extrêmes. Par ailleurs, l'accent porte davantage sur le volet non-paramétrique. FRAISE se démarque également de CERISE car il a pour ambition d'aller jusqu'aux études d'impact avec un consortium ajusté en conséquence.

Ces projets, finançant exclusivement du fonctionnement, ont été renforcés par l'obtention du

financement d'un post-doctorat pour 12 mois (octobre 2017-octobre 2018) par le LabEx Numev. Ce projet s'intitulait *Simulation de processus spatio-temporels intégrant des extrêmes pour mesurer le risque inondation : approches semi et non-paramétriques*. Fátima Palacios-Rodriguez, la post-doctorante a ensuite pu poursuivre pendant une année, toujours sous ma direction, sur une problématique en parfaite continuité avec ce premier post-doctorat grâce à un financement de l'Inria. Le travail mené dans ce cadre est une des contributions présentées dans le chapitre 3 de ce document. Lors de cette collaboration, nous avons notamment généralisé le travail **[G6]** proposant une méthode de simulations semi-paramétrique d'événements extrêmes spatio-temporels s'inscrivant dans le paradigme des processus de Pareto **[G16, G20]**.

J'ai également fait de ces **questions de recherche sur la modélisation et la simulation d'évènements extrêmes spatio-temporels un pilier du projet de recherche de l'équipe projet Inria LEMON** dont je suis membre depuis le 1er septembre 2017 (voir aussi **[G14]** traitant de l'intérêt des simulations stochastiques de pluies pour l'étude des inondations urbaines). Cette équipe-projet Inria, basée sur Montpellier et dépendant de Inria Nice Sophia-Antipolis, impliquant deux UMR que sont l'Intitut Montpellierrain Alexander Grothendieck (IMAG) et HydroSciences Montpellier (HSM) a été officiellement créée en janvier 2019. Cette nouvelle équipe se compose de 1 chercheur et 4 enseignants-chercheurs (dont 1 IMAG et 3 HSM). LEMON est une équipe interdisciplinaire travaillant sur le développement, l'analyse et l'application de modèles déterministes et stochastiques - éventuellement couplés - pour des processus littoraux. Les outils mathématiques utilisés dans cette équipe sont à la fois déterministes, probabilistes et statistiques. Plus précisément, les applications vont de l'océanographie régionale à la gestion côtière, y compris l'évaluation des risques naturels littoraux (submersion et inondations urbaines, tsunamis, pollution).

Cette recherche que je mène pour l'étude des extrêmes spatiaux s'accompagne en effet d'une réelle volonté de me nourrir des applications en science de l'environnement. En particulier l'interprétation physique des modèles et des paramètres éventuels associés est pour moi incontournable dans la phase de proposition de ces derniers. Les données sur lesquelles j'ai principalement travaillé sont des données de pollution **[G8, G10, G15]**, des hauteurs de vague dans l'étude du risque côtier **[G6, G18, G19]** et des précipitations **[G1, G3, G7, G14, G16, G17, G20]**. Les précipitations sont l'un des processus climatiques les plus complexes du fait de sa nature binaire (présence/absence), de l'importance de l'agrégation de valeurs fortes sur l'espace et le temps et des fortes variations pouvant apparaitre même à de très petites échelles spatiales et temporelles.

Ces travaux ont été menés dans le cadre de collaborations. J'ai en effet pu au cours de mon parcours développer de **nombreuses collaborations tant sur les aspects théoriques que sur les aspects appliqués**. Sur les aspects appliqués j'ai pu développer un très fort réseau local. J'ai en effet instauré un lien étroit avec des membres de Géosciences à Montpellier sur les aspects littoraux mais aussi et surtout une collaboration très forte avec des membres d'HydroSciences Montpellier dont certains sont également membres de LEMON.

Dans un tout autre contexte, en lien avec Sanofi, je me suis intéressée à un autre champ de recherche, en biostatistique médicale (voir aussi **[G13]**). Il s'agissait de proposer une méthodologie statistique pour le problème de la sélection des doses en développement clinique sous l'angle de la théorie de la décision et des fonctions d'utilité. Dans la suite de ce manuscrit, j'ai fait le choix de ne pas détailler cet axe et renvoie le lecteur intéressé aux articles correspondants **[G2, G4, G5, G21, G22]**.

En effet, dans la partie II de ce document qui propose notamment une analyse d'une partie des travaux scientifiques passés, j'ai choisi de ne présenter qu'une partie de ma recherche récente. M'intéressant particulièrement à l'étude des dynamiques temporelles et spatiales dans les extrêmes, je propose dans le chapitre 2 de présenter ma contribution portant sur la prise en compte de l'indépendance asymptotique dans des modèles spatiaux et spatio-temporels. Une autre de mes contributions principales ces dernières années porte sur la simulation c'est-à-dire la génération d'épisodes extrêmes pour étudier leur impact. Cela fait l'objet du chapitre 3.

Les objectifs et les méthodes de ces deux principaux chapitres présentent des différences et des similitudes. Alors que le chapitre 2 répond à un objectif de modélisation, le chapitre 3 propose une méthode de simulation d'épisodes extrêmes, basée sur une technique d'amplification, et répond donc au verrou numéro 3 identifié dans CERISE présenté ci-dessus. Une approche non-paramétrique pour la structure de dépendance y est considérée contrairement au chapitre 2 où des approches paramétriques sont exploitées. Le chapitre 3 fait l'hypothèse de dépendance asymptotique alors que le chapitre 2 lève quant à lui le premier verrou proposant une modélisation concernant l'indépendance asymptotique. Pour ce qui est de la prise en compte d'une dépendance temporelle (second verrou de CERISE), les deux chapitres apportent des contributions compte-tenu des modélisations spatio-temporelles qui y sont présentées.

Enfin, la partie II se termine avec le chapitre 4 dédié à la présentation de mon projet de recherche. La bibliographie est ensuite présentée, à l'exception de mes contributions toutes rassemblées et présentées dans la partie précédente (Partie I).

Les annexes constituent la partie III qui contient l'intégralité des 5 articles **[G1, G3, G6, G7, G20]**.

Chapitre 2

Prise en compte de l'indépendance asymptotique dans des modèles spatiaux ou spatio-temporels pour les extrêmes

Sommaire

2.1 Introduction

Dans un cadre spatial, les processus max-stables (Smith, 1990; Schlather, 2002; de Haan, 1984; Davison *et al.* , 2012; Davison & Gholamrezaee, 2012; Opitz, 2013) apparaissent comme les

modèles limites naturels de maxima spatiaux et leur utilisation est aujourd'hui courante pour les applications. Ces modèles sont basés sur une hypothèse forte de dépendance asymptotique. En d'autres termes, ils supposent que la dépendance reste la même pour tout niveau extrême du phénomène considéré. Ces modèles ne sont pas adaptés à des situations où la dépendance extrémale diminue pour des valeurs devenant extrêmement élevées, voire disparait asymptotiquement. Cette situation (indépendance asymptotique) semble pourtant être caractéristique de certaines données telles que les pluies (Davison *et al.* , 2013; Thibaud *et al.* , 2013; Le *et al.* , 2018). La négliger peut conduire à des interprétations erronées, typiquement à un biais dans l'évaluation de risques associés. À titre d'exemple Le *et al.* (2018) montrent qu'en cas d'indépendance asymptotique, le facteur de réduction surfacique (ARF - *Areal Reduction Factor*), important dans l'étude du risque inondation, diminue quand la période de retour augmente, contrairement au cas de dépendance asymptotique pour lequel il reste constant. Dans la littérature, peu de modèles spatiaux, et a fortiori spatio-temporels, dédiés aux extrêmes ont été proposés dans un contexte d'indépendance asymptotique. Il s'agit d'un champ de recherche très récent. Dans un cadre spatial, certains modèles asymptotiquement indépendants notamment basés sur les modèles inverses max-stables ainsi que des modèles de max-mélange, ont été proposés par Wadsworth & Tawn (2012). Partant de ces modèles, nous avons spécifié dans une collaboration avec J.N. Bacro et C. Gaetan **[G7]** une configuration d'un modèle de max-mélange, avons étudié ses propriétés et l'avons mis en œuvre sur des données de précipitations en Australie. Ce modèle hybride est extrêmement flexible et c'est ce qui en fait sa force. Il permet une situation de dépendance asymptotique pour des sites peu éloignés puis d'indépendance asymptotique pour des sites à distances intermédiaires et enfin un cadre indépendant pour de très longues distances. Il s'agit d'une configuration suggérée par l'étude exploratoire de données de pluie en Australie et notamment par les résultats d'estimations des coefficients de dépendance de queue calculés pour des paires de sites à différentes distances. En effet, sur cette zone, les graphiques associés qui seront présentés en section 2.3.1, suggèrent que même si les données présentent de l'indépendance asymptotique, l'intérêt de la prendre en compte dans la modélisation ne concernerait que les distances suffisamment grandes (>500kms). Ce genre de considération semble néanmoins dépendre d'un certain nombre d'éléments (zone géographique, type de pluies...). Pour des données horaires et dans la région de France méditerranéenne étudiée dans **[G3]** (voir aussi Figure 2.4), l'indépendance asymptotique semble vérifiée y compris pour des distances relativement petites. De plus, les estimations des coefficients de dépendance de queue mis en œuvre dans un cadre temporel suggèrent également une indépendance asymptotique dans le temps pour de petits écarts temporels. Dans un travail en collaboration avec J.N. Bacro, C. Gaetan et T. Opitz, nous avons proposé un second modèle **[G3]** permettant alors une indépendance asymptotique dans le temps et dans l'espace (et stricte indépendance pour des distances et écarts temporels suffisamment grands). Le modèle proposé est un modèle spatio-temporel. Une autre de ses forces est qu'il modélise des dépassements et se prête alors à une interprétation à l'échelle de l'événement, ce qui est fondamental pour certaines applications pratiques. Notre approche est hiérarchique et est basée sur la représentation de la distribution de Pareto généralisée comme mélange d'une distribution exponentielle avec une loi Gamma, ce qui permet d'être cohérents avec la théorie univariée des extrêmes. Nous nous sommes appuyés sur le travail de Bortot & Gaetan (2014) qui partent du même constat et l'exploitent dans un cadre temporel. Notre processus latent peut également être considéré comme une version spatio-temporelle des processus *trawl* (Barndorff-Nielsen *et al.* , 2014) exploités pour des valeurs extrêmes par Noven *et al.* (2018) dans un cadre temporel. L'utilisation de modèles hiérarchiques pour les extrêmes spatiaux ou spatio-temporels n'est pas nouvelle et nous pouvons, par exemple, citer les travaux de Casson & Coles (1999); Cooley *et al.* (2007); Gaetan & Grigoletto (2007) ou encore Sang & Gelfand (2009). Notre approche, contrairement à certaines autres approches

hiérarchiques proposées, se démarque par le fait qu'elle n'est pas basée sur des distributions gaussiennes et que les marginales sont naturellement en accord avec la théorie des extrêmes univariée (par opposition à être des mélanges de distributions des extrêmes).

L'objectif de ce chapitre est de présenter ma contribution à l'élaboration de modèles permettant la prise en compte de l'indépendance asymptotique. Après avoir rappelé quelques définitions formalisant le concept de (in)dépendance asymptotique en section 2.2, je présente dans ce chapitre l'élaboration de deux modèles permettant la prise en compte de l'indépendance asymptotique, chacun faisant l'objet d'une section du chapitre (voir sections 2.3 et 2.4). Ces deux articles se trouvent en annexe de ce document partie III. Dans les deux cas, il s'agit de travaux qui vont de la proposition d'un modèle probabiliste et de son étude théorique en terme de dépendance extrémale jusqu'à la mise en application avec interprétation sur des données complexes de précipitation. Enfin je conclus ce chapitre par quelques éléments de travaux en cours et perspectives sur cette thématique.

Les travaux auxquels j'ai contribué sur cette thématique ont donné lieu à une revue **[G19]** et à deux articles dans des journaux internationaux, à savoir JASA et JSPI **[G3, G7]**. J'ai également pu les présenter personnellement lors de 5 conférences internationales invitées **[C3, C4, C5, C7, C8]** et 2 conférences internationales **[C15, C16]**. J'ai par ailleurs présenté ce travail dans un certain nombre de séminaires comme par exemple au CMAP à Polytechnique dans le cadre de la chaire "Stress Testing" en novembre 2019. J'ai également présenté les modèles et résultats obtenus à une large communauté scientifique incluant des experts en hydrologie pour valider et diffuser ces nouvelles approches (présentation aux journées scientifiques du LabEx Numev, au séminaire d'HydroSciences Montpellier, INRAE d'Aix en Provence).

2.2 Notion d'indépendance asymptotique et indicateurs bivariés associés

2.2.1 Vers l'indépendance asymptotique

Soient $(X_1, Y_1), ..., (X_n, Y_n)$ des vecteurs aléatoires indépendants de distribution jointe commune F et $\mathbf{M}_n = (\max_{1\leq i\leq n} X_i, \max_{1\leq i\leq n} Y_i)$ le vecteur de maxima. S'il existe des suites de vecteurs $\mathbf{a_n} > 0$ et $\mathbf{b_n}$ telles que

$$\mathbb{P}(\mathbf{M_n} \leq \mathbf{a_n}\,\mathbf{z} + \mathbf{b_n}) = F^n(\mathbf{a_n}\,\mathbf{z} + \mathbf{b_n}) \longrightarrow G(\mathbf{z}) \tag{2.1}$$

quand $n \to \infty$ avec $\mathbf{z} \in \mathbb{R}^2$ et G une distribution non dégénérée, alors G est une distribution des valeurs extrêmes et on dit que F appartient au domaine d'attraction de G ($F \in D(G)$). Il est important de noter que la distribution G est alors max-stable, c'est-à-dire que pour tout k positif, il existe des suites $a_k > 0$ et b_k telles que $G^k(a_k\mathbf{z} + b_k) = G(\mathbf{z})$. La distribution G est à la fois caractérisée par ses deux marginales et par sa structure de dépendance. De la théorie des valeurs extrêmes univariée, on sait que les marginales sont distribuées selon la loi des valeurs extrêmes généralisées (GEV). Sans perte de généralité, on peut supposer une GEV particulière (Resnick, 1987) et un choix classique consiste à considérer la loi de Fréchet standard $\Psi(y) = \exp(-y^{-1})$, $y > 0$. Le coefficient extrémal $\theta \in [1, 2]$ se définit de la façon suivante $G(z, z) = \Psi^\theta(z)$. Dans le cas particulier où G correspond au produit des marges ($\theta = 2$), on dit

que X et Y sont **asymptotiquement indépendants** et dans ce cas, il est bien connu qu'une approche max-stable devient inopérante pour tout calcul de probabilités d'événements extrêmes.

Des modèles alternatifs multivariés basés sur la modélisation de la probabilité de survie ont alors été proposés par Ledford & Tawn (1996, 1997). En posant (X_P, Y_P) un vecteur bivarié avec marges Pareto standard (de fonction de répartition $H(x) = 1 - 1/x$, $x \geq 1$), ils proposent, pour $n \to \infty$, la caractérisation suivante :

$$\mathbb{P}\left(X_P > nx, Y_P > ny\right) \quad \sim \quad n^{-1/\eta} x^{-c_1} y^{-c_2} \mathcal{L}(nx, ny) \tag{2.2}$$

où $c_1 + c_2 = \frac{1}{\eta}$, $\mathcal{L}$ une fonction bivariée à variation lente et $0 < \eta \leq 1$. Le paramètre η, appelé coefficient de dépendance de queue, est une constante qui détermine la vitesse de décroissance de la fonction de survie jointe pour de grandes valeurs. Partant de 2.2 et supposant $z = nx = ny$, nous retrouvons le modèle de Ledford & Tawn (1996)

$$\mathbb{P}\left(X_P > z, Y_P > z\right) \quad \sim \quad z^{-1/\eta} \mathcal{L}(z) \tag{2.3}$$

où $\mathcal{L}(z)$ est une fonction à variation lente univariée.

Repartant de (2.2), pour élargir les axes d'extrapolation possibles et considérer des cas aussi larges que possibles de dépendance, un nouveau modèle a été présenté dans la thèse de N. Dalhoumi que j'ai pu co-encadrer (soutenue à Montpellier le 25/09/2017).

Soit (X_P, Y_P) un vecteur bivarié avec des distributions marginales Pareto standard. On suppose que pour $\beta > 0$, $\gamma > 0$ et $(x, y) \in [1, \infty)^2$:

$$\mathbb{P}(X_P > n^\beta x, Y_P > n^\gamma y) = \mathcal{L}(n^\beta x, n^\gamma y) n^{-\kappa(\beta,\gamma)} x^{\frac{-\kappa(\beta,\gamma)}{2\beta}} y^{\frac{-\kappa(\beta,\gamma)}{2\gamma}} \tag{2.4}$$

où κ est la fonction définie dans le modèle de Wadsworth & Tawn (2013) et $\mathcal{L}$ est une fonction bivariée vérifiant

$$\lim_{\min(n^\beta, n^\gamma) \to \infty} \frac{\mathcal{L}(n^\beta x, n^\gamma y)}{\mathcal{L}(n^\beta, n^\gamma)} = \lim_{n \to \infty} \frac{\mathcal{L}(n^\beta x, n^\gamma y)}{\mathcal{L}(n^\beta, n^\gamma)} = g_{(\beta,\gamma)}(x, y) \tag{2.5}$$

avec la fonction g vérifiant la propriété d'homogénéité d'ordre zéro non standard suivante : pour tout $(\beta, \gamma) \in \mathbb{R}_+^2 \backslash \{\mathbf{0}\}$ et $c > 0$

$$g_{(\beta,\gamma)}(c^\beta x, c^\gamma y) = g_{(\beta,\gamma)}(x, y). \tag{2.6}$$

Cette modélisation convient à la fois à des cas de dépendance asymptotique et d'indépendance asymptotique. D'un point de vue théorique, cette approche peut se voir comme une généralisation de Wadsworth & Tawn (2013) dans le sens où les axes d'extrapolation, qui ont comme dans Wadsworth & Tawn (2013) des angles variables (dans le cas de marges exponentielles), peuvent ne pas passer par l'origine. En pratique cet apport reste limité en dehors d'un cadre de simulation. En effet, la difficulté d'identifier les paramètres associés à cette flexibilité rend le modèle proposé difficilement exploitable en pratique.

2.2.2 Indicateurs bivariés de la dépendance extrémale

C'est en se basant sur le comportement de queue de la probabilité conditionnelle $\mathbb{P}(X > F_X^{\leftarrow}(q) | Y > F_Y^{\leftarrow}(q))$ quand q tend vers 1, où $F_X^{\leftarrow}$, $F_Y^{\leftarrow}$ sont les distributions inverses généralisées

de X et de Y (Sibuya, 1960; Coles *et al.* , 1999) qu'il est possible d'explorer la dépendance extrémale dans un vecteur aléatoire bivarié (X, Y). La limite obtenue quand $q \to 1^-$ définit le coefficient de queue χ comme suit :

$$\chi(q) := \frac{\mathbb{P}(X > F_X^{\leftarrow}(q), Y > F_Y^{\leftarrow}(q))}{\mathbb{P}(Y > F_Y^{\leftarrow}(q))} \to \chi, \qquad q \to 1^-.$$

Le vecteur aléatoire (X, Y) est dit asymptotiquement dépendant si χ est strictement positif et le cas $\chi = 0$ correspond à la situation d'indépendance asymptotique (Sibuya, 1960). Dans le cas d'une distribution max-stable G comme définie en (2.1), il est bien connu que ce coefficient de queue χ est égal à $2 - \theta$ avec $\theta \in [1, 2]$ le coefficient extrémal.

Pour obtenir une caractérisation plus fine de la vitesse de décroissance de queue bivariée sous l'indépendance asymptotique, plus rapide que la vitesse de décroissance marginale, Coles *et al.* (1999) ont introduit le paramètre $\overline{\chi}$ défini comme la limite de la quantité $\bar{\chi}(q)$ définie ci-dessous.

$$\bar{\chi}(q) := \frac{2 \log \mathbb{P}(Y > F_Y^{\leftarrow}(q))}{\log \mathbb{P}(X > F_X^{\leftarrow}(q), Y > F_Y^{\leftarrow}(q))} - 1 \to \bar{\chi} \in (-1, 1], \qquad q \to 1^-.$$

Plus $|\bar{\chi}|$ est grand, plus la dépendance est forte, le cas $|\bar{\chi}| = 0$ correspondant à la quasi-indépendance (*near independence*). Sous le modèle (2.3), $\bar{\chi}$ et η sont reliés par la relation $\bar{\chi} = 2\eta - 1$.

2.3 Modèle spatial de max-mélange

2.3.1 Présentation des données et motivation

Des résumés de la dépendance extrémale basés sur l'étude des paires de variables aléatoires ont été proposés (Coles *et al.* , 1999) comme rappelé en section 2.2.2. Pour un processus spatial stationnaire $Z = \{Z(s), s \in \mathcal{D} \subset \mathbb{R}^2\}$ de distribution marginale F, la dépendance extrémale entre deux sites s et $s + h$ peut être caractérisée au moyen de la fonction

$$\chi(h) = \lim_{q \to 1^-} \chi(h, q) = \lim_{q \to 1^-} \mathbb{P}(F(Z(s + h)) > q \mid F(Z(s)) > q). \tag{2.7}$$

La fonction $\chi(h, \cdot)$ peut être interprétée comme une mesure de la force de dépendance liée à un niveau de quantile qui est présente entre 2 sites s et $s + h$. Le cas $\chi(h) = 0$ correspond à une situation d'indépendance asymptotique par opposition au cas $\chi(h) \neq 0$ de dépendance asymptotique.

Sous l'indépendance asymptotique, la fonction $\chi(h)$ est d'un intérêt limité. En supposant de nouveau que Z est un processus spatial stationnaire de marginale F, la quantité $\bar{\chi}(h, q)$ se définit de la façon suivante

$$\bar{\chi}(h, q) = \frac{2 \log \mathbb{P}(F(Z(s)) > q)}{\log \mathbb{P}(F(Z(s)) > q, F(Z(s + h)) > q)} - 1, \qquad 0 \leq q \leq 1. \tag{2.8}$$

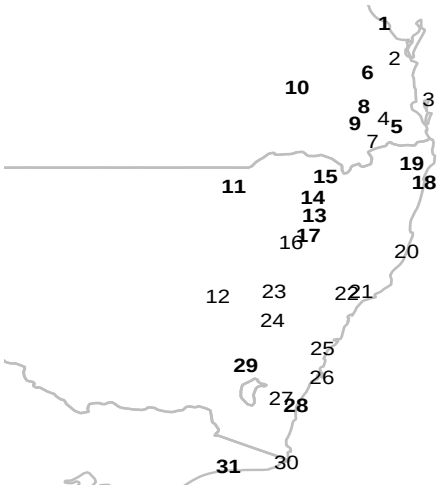


FIGURE 2.1 – Localisations géographiques des 31 stations météorologiques dans l'est de l'Australie. Les stations en gras ont été utilisées pour l'inférence alors que les autres sont les stations de validation.

La limite $\bar{\chi}(h) = \lim_{q \to 1^-} \bar{\chi}(h,q)$, avec $-1 < \bar{\chi}(h) \leq 1$, fournit une mesure qui augmente avec la dépendance extrémale entre $Z(s)$ et $Z(s+h)$ (Coles *et al.* , 1999), le cas $\bar{\chi}(h) = 1$ correspondant à la dépendance asymptotique.

Nous avons utilisé ces résumés de dépendance sur des cumuls journaliers de précipitations hivernales, enregistrées dans 31 sites dans l'est de l'Australie (Figure 2.1). Les estimations non paramétriques de $\chi(h,q)$ et de $\overline{\chi}(h,q)$ se trouvent en Figure 2.2. On peut noter qu'identifier la situation entre dépendance asymptotique et indépendance asymptotique au moyen des estimations de $\chi(h,q)$ ou de sa version limite $\chi(h)$ lorsque $q \to 1^-$, n'est pas facile en pratique, notamment pour les extrêmes de pluie (Serinaldi *et al.* , 2014). Néanmoins, dans le cas présent (Figure 2.2), quand q tend vers 1, $\chi(h,q)$ décroit et semble s'écraser vers 0 au niveau d'un coude autour de 500 kms voire un peu avant. Cela suggère que la dépendance asymptotique est présente jusqu'à une distance $r_1 = 500$ kms puis que l'indépendance asymptotique prévaut pour les distances supérieures à r_1. Enfin en regardant également le graphique de droite toujours sur la Figure 2.2, il n'est pas déraisonnable d'imaginer être en situation d'indépendance exacte pour les plus grandes distances ($> r_2 = 1000$ kms environ).

Le modèle considéré dans cette section est un modèle de max-mélange à deux composantes défini en tout point comme le maximum entre deux modèles particuliers pondérés (voir également Wadsworth & Tawn (2012)). Un modèle max-stable, et donc asymptotiquement dépendant (sauf cas indépendance stricte), constituera la première composante alors qu'un modèle asymptotiquement indépendant sera considéré pour la seconde. Les choix faits pour ces deux modèles constituant le mélange permet d'étendre les cadres d'applications de Wadsworth & Tawn (2012). Des précisions sur les choix faits sont données en sous-section suivante alors que le modèle de mélange est quant à lui présenté en section 2.3.3.

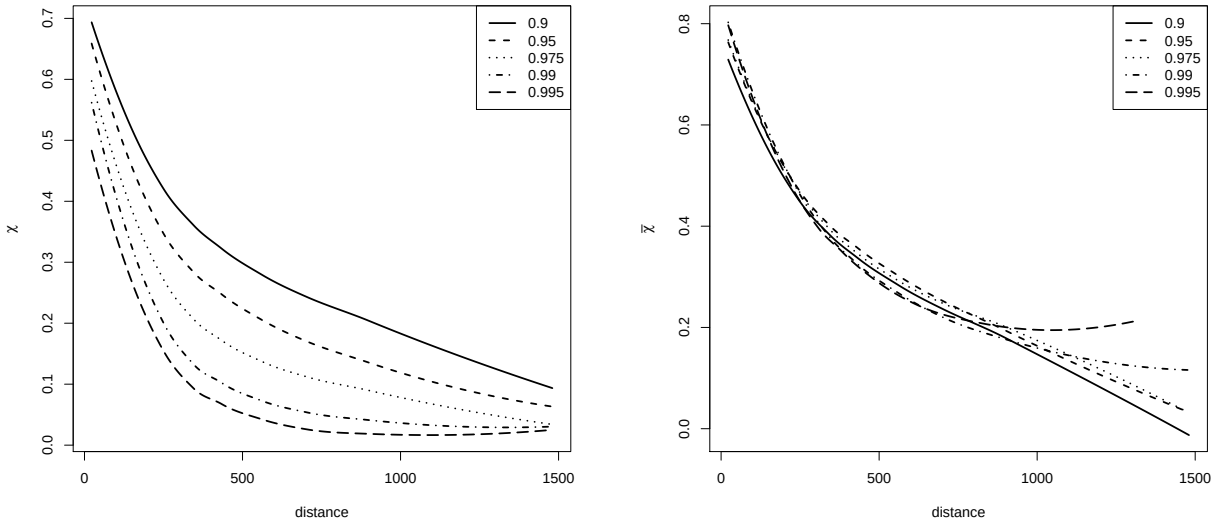


FIGURE 2.2 – Valeurs lissées des estimations empiriques des fonctions $\chi(h,q)$ (gauche) et $\bar{\chi}(h,q)$ (droite) pour différentes valeurs de q.

2.3.2 Extrêmes spatiaux

Modèles pour la dépendance asymptotique

Les processus max-stables (de Haan, 1984) sont une généralisation en dimension infinie de la théorie des valeurs extrêmes multivariée. Le processus stochastique $X = \{X(s), s \in \mathcal{D}\}$, où $\mathcal{D}$ est un domaine spatial, est un processus max-stable si et seulement si il existe des fonctions $a_n(\cdot) > 0$ et $b_n(\cdot)$ sur $\mathbb{R}$ telles que

$$\max_{1 \leq i \leq n} \frac{X_i(s) - b_n(s)}{a_n(s)} \stackrel{d}{=} X(s) \tag{2.9}$$

où $X_1, X_2, \ldots$ sont des copies indépendantes de X. Dans la suite de la présentation de ce travail et sans perte de généralité, $\mathcal{D}$ est un sous-ensemble de $\mathbb{R}^2$ et on supposera que le processus max-stable a des marginales Fréchet standard i.e. $\mathbb{P}(X(s) \leq x) = \Psi(x) = \exp(-1/x)$, $x > 0$.

Un processus max-stable a une représentation spectrale (de Haan, 1984; Schlather, 2002). On va supposer que r_i, $i \geq 1$, sont les points d'un processus de Poisson sur $(0, \infty)$ d'intensité dr. On suppose également que W_i, $i \geq 1$, sont des copies indépendantes et identiquement distribuées (i.i.d.) d'une fonction aléatoire continue à valeurs réelles $W = \{W(s), s \in \mathcal{D}\}$, indépendante des $\{r_i\}$ et telle que $\mathbb{E}[W^+(s)] = \mu \in (0, \infty)$, où $W^+(s) = \max(W(s), 0)$. Alors

$$X(s) = \mu^{-1} \max_{i \geq 1} W_i^+(s)/r_i \tag{2.10}$$

est un processus max-stable de marginales Fréchet standard.

Le choix d'une expression particulière pour W_i conduit à des exemples connus de processus max-stables. Parmi les plus connus, nous pouvons citer le processus de Smith (1990) basé

sur la densité gaussienne (*the Gaussian extreme value process*), le processus gaussien extrémal (Schlather, 2002) ou encore le processus Brown-Resnick (Kabluchko *et al.* , 2009) et le processus t-extremal (Opitz, 2013). Dans la suite, nous nous concentrons sur le processus gaussien extrémal tronqué (en anglais Truncated Extremal Gaussian - TEG). Le processus TEG a été introduit par Schlather (2002) et a été illustré par Davison & Gholamrezaee (2012). Comme dans le modèle extrémal gaussien (Schlather, 2002), le TEG est basé sur un processus gaussien et il est censuré sur un ensemble aléatoire compact. Plus précisément, on considère $W_i(s) = c\max(0, \varepsilon_i(s))\mathbb{1}_{B_i}(s-U_i)$ avec ε_i des copies indépendantes d'un processus gaussien stationnaire $\varepsilon = \{\varepsilon(s), s \in \mathcal{D}\}$ de moyenne nulle, de variance égale à 1 et de fonction de corrélation $\rho(\cdot)$, $\mathbb{1}_B$ est la fonction indicatrice avec $B \subset \mathcal{D}$ un ensemble aléatoire compact, dont on considère des copies indépendantes B_i et U_i sont les points d'un processus de Poisson homogène d'intensité 1 sur $\mathcal{D}$, indépendants des ε_i.

En choisissant la constante c telle que $c^{-1} = \mathbb{E}\left(\max\{W_i(s), 0\}\mathbb{1}_{B_i}(x - U_i)\right)$ et en considérant r_i, $i \geq 1$ comme précédemment, le processus TEG est défini comme

$$X(s) = \max_{i\geq 1} \frac{W_i(s)}{r_i}. \tag{2.11}$$

La distribution marginale de X est Fréchet standard et la distribution bivariée est donnée par

$$\mathbb{P}\left(X(s) \leq t_1, X(s+h) \leq t_2\right) = \exp\left\{-\left(\tfrac{1}{t_1} + \tfrac{1}{t_2}\right)\left[1 - \tfrac{\alpha(h)}{2}\left(1 - \left(1 - 2\tfrac{(\rho(h)+1)t_1t_2}{(t_1+t_2)^2}\right)^{1/2}\right)\right]\right\} \tag{2.12}$$

où $\alpha(h) = \mathbb{E}[|B \cap (h+B)|]/\mathbb{E}[|B|]$. Si B est un disque de rayon fixé r, nous obtenons l'approximation suivante pour $\alpha(h)$

$$\alpha(h) \simeq (1 - ||h||/(2r))\, \mathbb{1}_{[0,2r]}(||h||) \tag{2.13}$$

avec $||.||$ la distance euclidienne.

De nouveau, nous renvoyons le lecteur à Schlather (2002); Davison & Gholamrezaee (2012) pour plus de détails.

Généralisant le coefficient extrémal rappelé en section 2.2.2 comme une fonction de l'écart entre 2 sites, la fonction coefficient extrémal (Schlather & Tawn, 2003) est une mesure de la dépendance spécifique pour les processus max-stables. Étant donnée une paire de sites s et $s+h$, la fonction coefficient extrémal $\theta(h)$ est définie comme

$$\mathbb{P}(X(s) \leq x, X(s+h) \leq x) = \mathbb{P}(X(s) \leq x)^{\theta(h)}.$$

Ici $1 \leq \theta(h) \leq 2$ et $\theta(h) = 1$ ou $\theta(h) = 2$ correspond à la dépendance parfaite ou à l'indépendance stricte, respectivement. De nouveau, il est possible de montrer que $\theta(h) = 2 - \chi(h)$.

Remarquons également que pour un processus max-stable, toute distribution bivariée est max-stable et alors, à h fixé, la fonction $\chi(h, .)$ est constante.

Les cas particuliers du processus de Smith (1990) ou du processus Brown-Resnick (Kabluchko *et al.* , 2009) autorisent des dépendances extrémales qui vont de la dépendance parfaite à l'indépendance stricte pourvu que la distance $\|h\|$ augmente à l'infini. En revanche, l'extrémal

gaussien non tronqué (Schlather, 2002) suppose une dépendance asymptotique quelle que soit la distance considérée. Le processus TEG a pour fonction de coefficient extrémal

$$\theta(h) = 2 - \alpha(h)\left\{1 - 2^{-1/2}[1-\rho(h)]^{1/2}\right\} \tag{2.14}$$

qui atteint la valeur $\theta(h) = 2$ pour $||h||$ suffisamment grand. Cette caractéristique qui assure, à une distance finie, l'indépendance stricte dans les extrêmes, est une caractéristique clé qui est exploitée plus tard dans le modèle max-mélange utilisé et qui justifie le fait que nous nous concentrons sur ce processus.

Modèles pour l'indépendance asymptotique

Un vecteur multivarié est asymptotiquement indépendant si et seulement si toutes les paires de ses composantes sont asymptotiquement indépendantes (de Oliveira, 1962). Par conséquent, si toutes les distributions bivariées d'un processus stochastique sont asymptotiquement indépendantes, le processus stochastique sera dit asymptotiquement indépendant.

On considère que $\{Y(s), s \in \mathcal{D}\}$ est un processus spatial stationnaire avec des marginales Fréchet standard. Nous plaçant dans le cadre du modèle de queue décrit en 2.3, nous supposerons que

$$\mathbb{P}(Y(s) > z, Y(s+h) > z) \sim z^{-1/\eta(h)}\mathcal{L}_h(z) \qquad \text{pour } z \to \infty \tag{2.15}$$

où $\mathcal{L}_h(\cdot)$ est une fonction à variation lente univariée.

La fonction $\eta(h)$ varie entre 0 et 1 et détermine la vitesse de décroissance de la probabilité de queue bivariée $\mathbb{P}(Y(s) > z, Y(s+h) > z)$ pour z grand. Le modèle relativement simple défini en (2.15) est en fait un modèle qui est très général pour la distribution de queue bivariée et qui peut fournir, comme détaillé ci-après, une mesure de la dépendance extrémale entre $Y(s)$ et $Y(s+h)$ via le coefficient $\eta(h)$. L'indépendance asymptotique correspond à $\eta(h) < 1$ et dans ce cas $\eta(h)$ mesure le degré de dépendance résiduelle pour deux sites séparés de h, où $\eta(h) > 1/2$ et $\eta(h) < 1/2$ indiquent une association respectivement positive et négative. Quand les variables $Y(s)$ et $Y(s+h)$ sont indépendantes, $\eta(h) = 1/2$. Dans la littérature, il existe quelques exemples de processus asymptotiquement indépendants. Je citerai ci-dessous simplement les processus gaussiens et les inverses max-stables.

Exemple 1 : processus gaussien stationnaire
Soit $\{Z(s), s \in \mathcal{D}\}$ un processus gaussien stationnaire d'espérance nulle, de variance égale à 1 et de fonction de corrélation $\rho(h)$. Comme deux variables gaussiennes non parfaitement corrélées sont asymptotiquement indépendantes (Sibuya, 1960), le processus spatial $Y(s) = -1/\log(\Phi(Z(s)))$ a des marges Fréchet standard et vérifie 2.15 avec $\eta(h) = \{1+\rho(h)\}/2 = \{1+\bar{\chi}(h)\}/2$.

Exemple 2 : processus inverse max-stable
Un processus inverse max-stable (Wadsworth & Tawn, 2012) est obtenu en prenant simplement l'inverse d'un processus max-stable. Plus précisément, si l'on considère un processus max-stable $\{X(s), s \in \mathcal{D}\}$ défini par (2.10), alors le processus

$$Y(s) = -1/\log[1 - \exp\{-1/X(s)\}]$$

est un processus asymptotiquement indépendant de marginales Fréchet. Pour tout h fixé, le coefficient de dépendance de queue est égal à $\eta(h) = 1/\theta(h)$ où $\theta(h)$ est le coefficient extrémal

du processus max-stable. Remarquons également que $\bar{\chi}(h,\cdot)$ est dans ce cas une constante. En d'autres termes, les distributions de survie bivariée d'un inverse max-stable sont uniquement liées aux fonctions de survie marginales du processus et ce quel que soit l'ordre de grandeur des événements extrêmes considérés.

2.3.3 Proposition du modèle max-mélange

Dans la suite, nous considérons que $X = \{X(s), s \in \mathcal{D}\}$ et $Y = \{Y(s), s \in \mathcal{D}\}$ sont deux processus spatiaux indépendants et stationnaires, de marges Fréchet standard et tels que X est un processus max-stable et Y un processus asymptotiquement indépendant. Le processus de max-mélange (MM) $Z = \{Z(s), s \in \mathcal{D}\}$ est alors défini de la façon suivante

$$Z(s) = \max(\beta X(s), (1-\beta)Y(s)), \qquad 0 \leq \beta \leq 1. \tag{2.16}$$

Le modèle MM a été introduit par Wadsworth & Tawn (2012) pour modéliser des situations pour lesquelles la dépendance extrémale pourrait varier avec la distance. Dans Wadsworth & Tawn (2012), diverses possibilités de processus max-stables avec leurs versions inversées comme processus asymptotiquement indépendants ont été considérées mais tous les modèles ajustés présentaient le même type de dépendance (dépendance asymptotique ou indépendance asymptotique) quelle que soit la distance considérée.

Pour étendre cette approche et répondre à notre objectif de modélisation de données semblant présenter 3 types de dépendance extrémale différentes en fonction des distances entre les sites, nous proposons dans cette contribution de considérer un modèle MM qui permette la dépendance asymptotique à des courtes distances, l'indépendance asymptotique à des distances intermédiaires et éventuellement l'indépendance à des distances plus importantes. Plus précisément, nous choisissons un processus TEG (2.11) comme processus max-stable X avec une fonction de covariance notée $\rho(\cdot)$, et nous élargissons la classe des processus asymptotiquement indépendants envisagés dans l'approche de Wadsworth & Tawn (2012) en considérant des processus asymptotiquement indépendants de marginales Fréchet standard dont les distributions bivariées vérifient le modèle de Ledford & Tawn (1996) (2.15) pour $\eta(h) < 1$.

Comme les deux processus X et Y sont indépendants, on trouve directement l'expression de la distribution bivariée pour une paire de sites

$$\begin{aligned}&\mathbb{P}(Z(s) \leq z_1, Z(s+h) \leq z_2) = \\ &\quad \exp\left\{-\beta\left(\tfrac{1}{z_1}+\tfrac{1}{z_2}\right)\left[1-\tfrac{\alpha(h)}{2}\left(1-\left(1-2\tfrac{(\rho(h)+1)z_1z_2}{(z_1+z_2)^2}\right)^{1/2}\right)\right]\right\} F_Y^h\left(\tfrac{z_1}{1-\beta}, \tfrac{z_2}{1-\beta}\right)\end{aligned} \tag{2.17}$$

où $F_Y^h(y_1,y_2) = \mathbb{P}(Y(s) \leq y_1, Y(s+h) \leq y_2)$. Comme $\mathbb{P}(Z(s) \leq z) = \mathbb{P}(Z(s) \leq z, Z(s+h) < \infty) = \exp(-1/z)$ le modèle a des marginales Fréchet standard.

2.3.4 Mesures de la dépendance extrémale associée au modèle

En exploitant la caractérisation (2.15), on obtient,

$$\mathbb{P}(Z(s) > z, Z(s+h) > z) = \frac{\beta(2-\theta(h))}{z} + \left(\frac{z}{1-\beta}\right)^{-1/\eta(h)} \mathcal{L}_h\left(\frac{z}{1-\beta}\right) + O(z^{-2}).$$

Par conséquent, il est possible de déduire que la fonction $\chi(h)$ s'exprime de la façon suivante

$$\chi(h) = \beta\,(2 - \theta(h)) = \beta\,\alpha(h)\left(1 - \sqrt{\frac{1-\rho(h)}{2}}\right).$$

En utilisant l'approximation (2.13), il découle que les paires de sites séparés par une distance $||h||$ sont asymptotiquement dépendantes si la distance est inférieure à $2r$ et asymptotiquement indépendantes sinon.

Si $2 - \theta(h) \neq 0$, nous pouvons conclure que $\bar{\chi}(h, q) \to 1$ quand $q \to 1$. Dans la situation où $2 - \theta(h) = 0$, nous montrons que $\bar{\chi}(h, q) \to 2\eta(h) - 1$ quand $q \to 1$. Se basant sur (2.13), nous pouvons résumer les résultats de la façon suivante

$$\bar{\chi}(h) = \mathbb{1}_{[0,2r)}(||h||) + (2\eta(h) - 1)\mathbb{1}_{[2r,\infty)}(||h||),$$

mettant en évidence les comportements différents selon les distances entre deux sites. Soit $R > 2r$ et supposant que $\eta(h) = 1/2$ pour $||h|| > R$, alors les paires de sites distants de $||h||$ sont asymptotiquement dépendantes pour $||h|| < 2r$, asymptotiquement indépendantes pour $2r \leq ||h|| \leq R$ et indépendantes pour $||h|| > R$. Par exemple, pour le processus gaussien stationnaire transformé de marges Fréchet standard et de corrélation $\rho_Y(h)$, nous avons

$$\bar{\chi}(h) = \mathbb{1}_{[0,2r)}(||h||) + \rho_Y(h)\mathbb{1}_{[2r,\infty)}(||h||).$$

Dans ce cas, si la fonction de corrélation $\rho_Y(\cdot)$ est telle que $\rho_Y(h) = 0$ quand $||h|| > R$, nous sommes pour les grandes distances en situation d'indépendance stricte.

2.3.5 Mise en œuvre du modèle

Pour estimer les paramètres du modèle, nous avons utilisé une approche par vraisemblance composite par paires (Lindsay, 1988; Varin, 2008). Padoan *et al.* (2010) ont utilisé cette approche dans le cadre de processus max-stables spatiaux s'appuyant sur les densités bivariées associées. Cette approche peut également s'utiliser à partir des densités bivariées des dépassements au-dessus d'un seuil élevé (Jeon & Smith, 2012; Wadsworth & Tawn, 2012; Bacro & Gaetan, 2014; Huser & Davison, 2014). En particulier, Wadsworth & Tawn (2012) proposent dans un cadre d'indépendance asymptotique de modéliser les dépassements au-delà d'un seuil suffisamment élevé en utilisant la densité du modèle pour peu qu'au moins l'une des deux composantes dépasse le seuil - et de censurer dans le cas contraire. Très récemment, Ahmed *et al.* (2017, 2019) ont travaillé sur ce modèle et se sont intéressés à d'autres méthodes d'estimation, notamment une approche par moindres carrés basée sur le F-madogramme.

Dans une étude de simulation (voir **[G7]** pour plus de détails), nous avons illustré la capacité de cette approche à estimer correctement les paramètres du modèle MM (2.16) en considérant un TEG comme processus max-stable et un processus gaussien, transformé pour avoir des marginales Fréchet standard, avec fonction de corrélation sphérique, comme processus asymptotiquement indépendant.

Il est également important de noter que dans le cas de l'utilisation d'une vraisemblance composite, Varin & Vidoni (2005) ont proposé un critère de sélection de modèle qui est en fait le

pendant du critère AIC dans le cadre de vraisemblance complète. Il s'agit du CLIC (Composite Likelihood Information Criterion). Dans une étude de simulations, également décrite en détail dans **[G7]**, l'identification du vrai modèle entre le modèle MM (2.16) et ses deux composantes par le CLIC donne des résultats particulièrement performants. Très récemment, également dans un contexte de vraisemblance composite, Abu-Awwad *et al.* (2019) proposent deux tests statistiques pour le paramètre de mélange $\beta \in (0, 1)$.

Le modèle proposé a également été mis en œuvre sur les données d'Australie présentées en section 2.3.1. Il s'agit de cumuls journaliers de pluies enregistrés entre 1955 et 2003 en 31 sites (Figure 2.1). L'étude exploratoire de la dépendance extrémale (voir 2.2) suggère, comme discuté en section 2.3.1, différents types de dépendance extrémale en fonction des distances entre les sites. Un modèle max-mélange tel que proposé semble tout à fait adapté, nous laissant également la possibilité de comparer différents modèles à l'aide du CLIC. Nous avons donc mené une étude assez complète en deux parties, la première avec une inférence sur les maxima annuels (période avril-septembre) et la seconde avec une inférence sur les dépassements. Trois catégories de modèles ont été comparées : le modèle MM et ses deux composantes prises seules. Autrement dit, nous avons cherché à comparer un modèle max-mélange, un modèle max-stable TEG et un modèle asymptotiquement indépendant. Pour la composante asymptotiquement indépendante, plusieurs modèles ont été considérés également : processus gaussien transformé avec deux types de fonctions de corrélation (exponentielle et sphérique) et processus inverse TEG. Les modèles ont été comparés à l'aide du CLIC.

Nous avons sélectionné un modèle de chaque catégorie et je reprends ici les notations de l'article **[G7]** par souci de cohérence. A_2 est un modèle max-mélange avec $\beta \neq 0$, B est un modèle TEG et C_3 un modèle asymptotiquement indépendant (plus précisément c'est un inverse TEG). Afin d'illustrer le comportement de ces modèles et de vérifier leur ajustement, nous avons considéré certaines probabilités conditionnelles (calculées empiriquement sur la base de simulations). Le site s_1 se trouve dans le coin supérieur droit de la carte (voir Figure 2.1) et est considéré comme le site de référence. De plus, nous considérons trois sous-ensembles de sites $\mathcal{S}_1 = \{s_2, s_3, s_6, s_8, s_{10}\}$, $\mathcal{S}_2 = \{s_{11}, s_{13}, s_{14}, s_{15}, s_{18}\}$ et $\mathcal{S}_3 = \{s_{25}, s_{26}, s_{27}, s_{28}, s_{29}\}$ qui correspondent à 3 classes de distances par rapport à s_1. Nous calculons alors les probabilités conditionnelles $\mathbb{P}(Z(s) > z, s \in \mathcal{S}_i \mid Z(s_1) > z)$, $i = 1, 2, 3$ pour différentes grandes valeurs de p telles que $\mathbb{P}(Z(s_1) \leq z) = p$ avec p compris entre 0.86 et 0.996. Quand les autres modèles sous-estiment ou surestiment les probabilités empiriques pour certains seuils et certaines classes de distance, l'ajustement du modèle MM A_2 (colonne 1 sur la Figure 2.3) donne de bons résultats pour différents seuils et toutes classes de distances. Ces résultats indiquent que le modèle de max-mélange tel que proposé ajoute une flexibilité de modélisation à l'analyse spatiale extrême et semble pouvoir englober différents degrés de dépendance spatiale extrême. De nouveau, le lecteur intéressé trouvera des compléments et précisions dans l'article correspondant **[G7]** également présent en annexe A partie III de ce document.

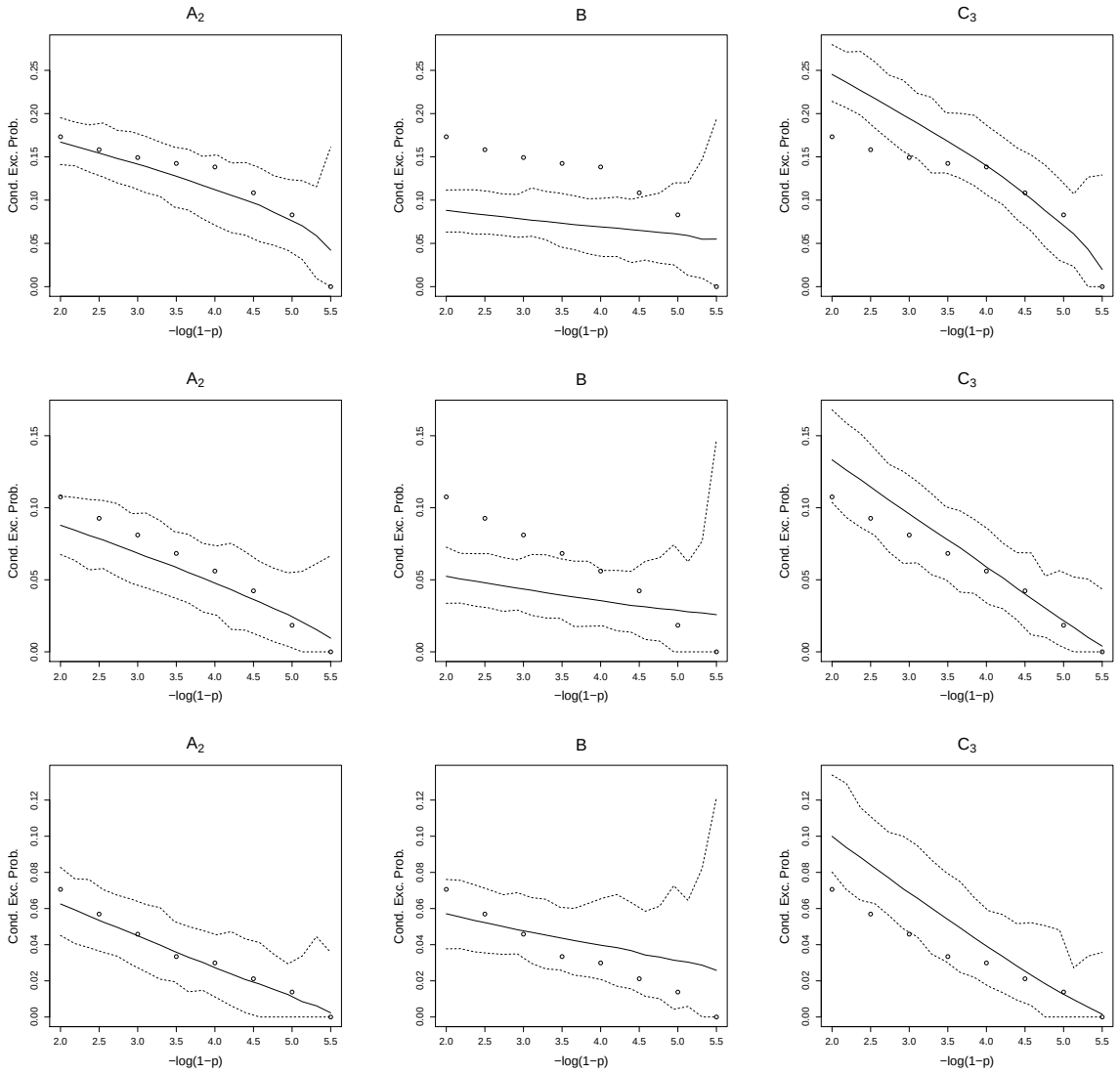


FIGURE 2.3 – Données journalières hivernales : valeurs empiriques et selon le modèle (avec paramètres estimés) des probabilités conditionnelles $\mathbb{P}(Z(s) > z, s \in \mathcal{S} \mid Z(s_1) > z)$. Les 3 colonnes correspondent aux modèles A_2 (max-mélange), B (max-stable) et C_3 (asymptotiquement indépendant). Ligne du haut : $\mathcal{S} = \mathcal{S}_1$ (sites proches) ; ligne du milieu $\mathcal{S} = \mathcal{S}_2$ (sites à distances intermédiaires) ; ligne du bas : $\mathcal{S} = \mathcal{S}_3$ (sites éloignés). Les valeurs $1-p$ sont telles que $\mathbb{P}(Z(s_1) > z) = 1 - p$.

2.4 Modèle asymptotiquement indépendant dans l'espace et dans le temps

2.4.1 Présentation des données et motivation

Dans cette section, nous nous intéressons à la modélisation de données horaires de précipitation mesurées en 50 stations dans le sud de la France (voir Figure 2.4). L'objectif est de proposer un modèle statistique spatio-temporel, physiquement interprétable, pour les dépassements, et qui soit capable de capturer les dépendances complexes et les dynamiques temporelles présentes dans les données.

Les résultats des analyses exploratoires de la dépendance extrémale présente dans ces données sont présentés en Figure 2.5. Plus précisément on y trouve les estimations de $\chi(q)$ et de $\overline{\chi}(q)$ pour des probabilités $q = 0.99, 0.995$. Cela a été effectué pour des paires en prenant simplement en compte une distance spatiale puis pour des paires en prenant cette fois-ci uniquement en compte un écart temporel. Les intervalles de confiance sont obtenus par une procédure bootstrap. Le fait que quand q augmente, $\chi(q)$ semble aller vers 0 indique que faire l'hypothèse d'une situation d'indépendance asymptotique est raisonnable que ce soit dans le cas spatial ou temporel. Par ailleurs les graphiques du $\overline{\chi}(q)$ font état d'une dépendance résiduelle strictement positive, du moins jusqu'à une certaine distance et un certain écart temporel.

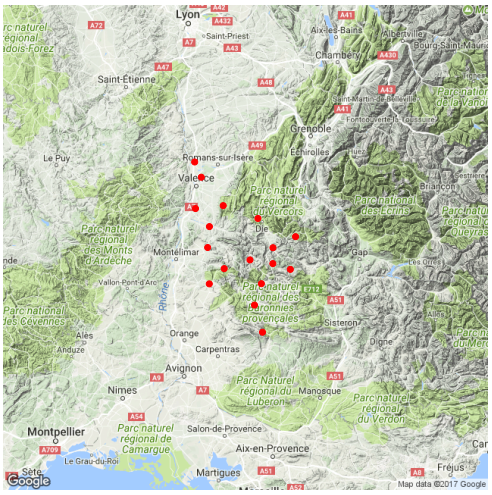


FIGURE 2.4 – Localisations géographiques des 50 stations météorologiques dans le sud-est français.

Quelques modèles pour extrêmes ont été proposés dans un cadre spatio-temporel. Sans être exhaustive, on peut citer les travaux de Davis *et al.* (2013a,b) qui étendent les processus max-stables de Brown-Resnick à un cadre spatio-temporel et proposent une inférence par vraisemblance par paires. Les processus max-stables TEG rappelés et utilisés dans la modélisation proposée en section 2.3 ont été généralisés dans une version spatio-temporelle par Huser & Davison (2014) et mis en œuvre sur des données horaires de précipitations. Ces derniers modélisent les tempêtes par des disques de rayon aléatoire, se déplaçant à une vitesse aléatoire pendant une durée aléatoire, s'appuyant donc sur des cylindres spatio-temporels. Le modèle proposé dans

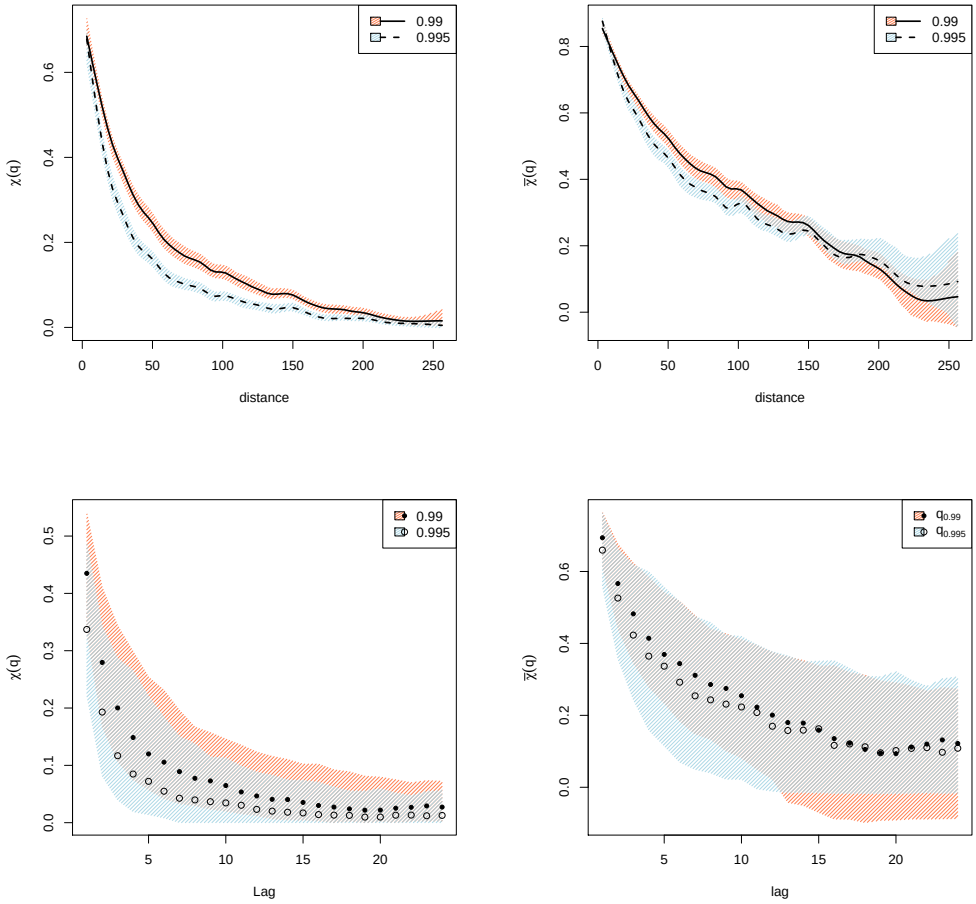


FIGURE 2.5 – Estimations empiriques de $\chi(q)$ (à gauche) et de $\bar{\chi}(q)$ (à droite) pour les données de précipitations. Les calculs sont faits sur des paires à une certaine distance spatiale (en haut) ou à une certaine distance temporelle (en bas). Les zones hachurées représentent des zones de confiance à 95% obtenues par une procédure bootstrap.

la suite s'appuie sur des objets géométriques similaires. Néanmoins, contrairement aux modèles max-stables, il s'agit d'un modèle asymptotiquement indépendant dans l'espace et dans le temps. Il est adapté aux dépassements au delà d'un seuil suffisamment élevé, ce qui permet d'une part d'exploiter plus d'informations des données et d'autre part de modéliser les données à l'échelle de l'événement lui-même. C'est aussi le cas des processus de Pareto dont il sera question dans le chapitre suivant qui sont le pendant des max-stables pour les dépassements et souffrent donc des mêmes limitations en terme de dépendance associée (dépendance asymptotique uniquement).

Dans cette section sera donc présentée la proposition d'un modèle hiérarchique spatio-temporel pour les dépassements au-dessus d'un seuil élevé qui permette l'indépendance asymptotique dans l'espace et dans le temps et qui est physiquement interprétable. Pour plus de détail, le lecteur est invité à consulter l'article correspondant publié dans JASA en 2019 **[G3]** et également présenté en annexe C partie III de ce document.

2.4.2 Proposition du modèle hiérarchique spatio-temporel pour des dépassements

Lorsque l'on s'intéresse aux dépassements d'une variable aléatoire X au delà d'un seuil élevé u, il est bien connu de la théorie des valeurs extrêmes que la distribution de Pareto généralisée (GPD pour *Generalized Pareto Distribution*) apparaissant comme la loi limite de la distribution conditionnelle des excès, est la distribution naturelle à considérer (voir notamment Pickands, 1975; Davison & Smith, 1990, pour les aspects probabilistes et statistiques). La fonction de répartition de la GPD est définie pour tout $y > 0$ par

$$GP(y; \sigma, \xi) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)_+^{-(1/\xi)}, \tag{2.18}$$

où $(a)_+ = \max(0, a)$, ξ est un paramètre de forme et σ un paramètre d'échelle positif. Le signe de ξ caractérise le domaine d'attraction de la distribution de X : $\xi > 0$ correspond au domaine d'attraction de Fréchet, $\xi = 0$ correspond au domaine de Gumbel et $\xi < 0$ correspond au domaine de Weibull.

Quand $\xi > 0$, la GPD peut s'exprimer comme un mélange d'une loi Gamma avec une loi exponentielle (Reiss & Thomas, 2007, p.157), i.e.,

$$V|\Lambda \sim \text{Exp}(\Lambda), \quad \Lambda \sim \text{Gamma}(1/\xi, \sigma/\xi) \quad \Rightarrow \quad V \sim GP(\,\cdot\,; \sigma, \xi), \tag{2.19}$$

où $\text{Exp}(b)$ correspond à la loi exponentielle de paramètre $b > 0$ et $\text{Gamma}(a, b)$ à la distribution Gamma avec un paramètre de forme $a > 0$ et $b > 0$. S'appuyant sur cette structure hiérarchique, nous développons une construction spatio-temporelle stationnaire pour modéliser les dépassements au-delà d'un seuil élevé qui possèdent des marges GPD pour les excès strictement positifs.

Première étape : structure spatio-temporelle hiérarchique. Nous considérons un processus aléatoire spatio-temporel stationnaire $Z = \{Z(x), x \in \mathcal{X}\}$ avec $x = (s, t)$ et $\mathcal{X} = \mathbb{R}^2 \times \mathbb{R}^+$, tel que s représente la localisation spatiale et t le temps. Sans perte de généralité, supposons que les marginales $Z(x)$ appartiennent au domaine d'attraction de Fréchet. Pour inférer le comportement de queue de $\{Z(x)\}$, on s'intéresse aux dépassements et plus précisément aux excès au delà d'un seuil élevé u

$$Y(x) = (Z(x) - u) \cdot \mathbf{1}_{(u,\infty)}(Z(x)). \tag{2.20}$$

D'après les résultats de la théorie des valeurs extrêmes, la GPD apparait comme un modèle naturel à considérer pour les valeurs $Y(x) > 0$. En suivant Bortot & Gaetan (2014), nous utilisons la représentation de la GPD comme un mélange d'une Gamma avec une exponentielle en supposant $\xi > 0$. En intégrant la dépendance spatio-temporelle dans un processus Gamma latent, le modèle proposé présente une dépendance spatio-temporelle à la fois dans les excès positifs $Z(x) - u > 0$ et dans le fait de dépasser ou non, c'est-à-dire dans $\mathbf{1}_{(u,\infty)}(Z(x))$.

En effet et plus formellement, on conditionne par un processus aléatoire spatio-temporel latent $\{\Lambda(x),\, x \in \mathcal{X}\}$ aux marginales $\Lambda(x) \sim \text{Gamma}(\alpha, \beta)$ et nous supposons

$$Y(x) \mid [\Lambda(x), Y(x) > 0] \sim \text{Exp}\left(\Lambda(x)\right), \tag{2.21a}$$

$$\mathbb{P}(Y(x) > 0 \mid \Lambda(x)) = e^{-\kappa \Lambda(x)}, \tag{2.21b}$$

où $\kappa > 0$ est un paramètre contrôlant le taux de dépassements du seuil.

Il est clair qu'une caractéristique importante du modèle ainsi construit est qu'il lie naturellement les probabilités de dépassements à l'amplitude des excès et par conséquent fournit une structure spatio-temporelle commune pour la partie positive et pour les zéros dans la distribution de $Y(x)$.

La distribution marginale de $Y(x)$ conditionnellement à $Z(x) > u$ correspond à la GPD, et la distribution (non-conditionnelle) de $Y(x)$ est

$$F(y;\sigma,\xi) = \begin{cases} p & \text{pour } y = 0, \\ p + (1-p)GP(y;\xi,\sigma) & \text{pour } y > 0, \end{cases} \tag{2.22}$$

avec un paramètre de forme $\xi = 1/\alpha$, un paramètre d'échelle $\sigma = (\kappa+\beta)/\alpha$, et $1-p$ la probabilité d'excéder u, i.e., $\mathbb{P}(Z(x) > u) = \mathbb{P}(Y(x) > 0) = 1 - p$. Cette probabilité de dépasser u,

$$\mathbb{P}(Z(x) > u) = \mathbb{E}\left(\mathbb{P}(Y(x) > 0|\Lambda(x))\right) = \mathbb{E}\left(e^{-\kappa\Lambda(x)}\right) = \left(\frac{\beta}{\kappa+\beta}\right)^{\alpha} \tag{2.23}$$

dépend de κ et correspond à la transformée de Laplace de $\Lambda(x)$ évaluée en κ.

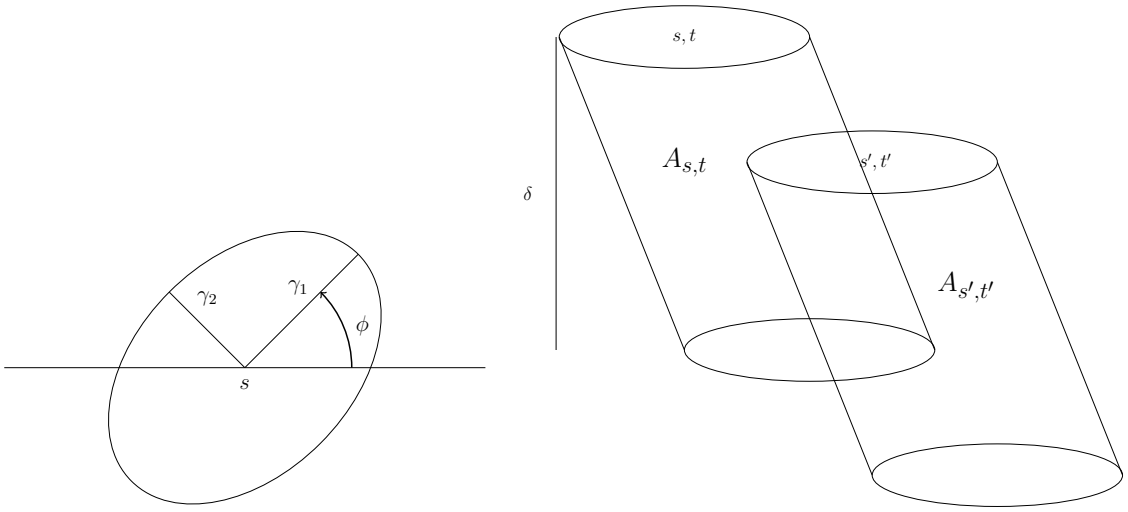
Lorsque l'on s'intéresse à des données de précipitations en région Méditerranéenne française, la condition $\xi > 0$ n'est pas restrictive car il s'agit d'un phénomène à queue lourde. Néanmoins il est tout à fait possible de relâcher cette hypothèse en transformant la distribution de $Y(x)$ (voir **[G3]** pour plus de détail).

Deuxième étape : dépendance spatio-temporelle avec des champs aléatoires Gamma
La dépendance spatio-temporelle est introduite au moyen d'un champ aléatoire stationnaire spatio-temporel noté $\{\Lambda(x), x \in \mathcal{X}\}$ aux marginales Gamma(α, β). En théorie, il est possible d'utiliser une large variété de modèles avec différents types de dépendance spatio-temporelle. Nous renvoyons le lecteur aux ouvrages de Cressie (1991); Cressie & Wikle (2011); Wikle *et al.* (2019) pour l'étude des statistiques spatiales et spatio-temporelles. À titre d'exemple, nous pourrions considérer un champ aléatoire gaussien spatio-temporel et en transformer les marginales pour respecter la contrainte. Cependant, nous avons privilégié une construction dans laquelle les distributions marginales Gamma apparaissent naturellement sans appliquer de transformations artificielles sur les marges.

En s'inspirant de Wolpert & Ickstadt (1998), nous avons considéré $\{\Lambda(x), x \in \mathcal{X}\}$ comme une convolution de processus aléatoire Gamma $\Gamma(dx)$, i.e., $\Lambda(x) = \int K(x,x')\Gamma(dx')$. Avant de nous intéresser au choix du noyau $K(x,x')$, définissons ci-après un champ aléatoire Gamma $\Gamma(dx)$ (Ferguson, 1973). Fixons $\mathcal{X} = \mathbb{R}^d$ et considérons $A \in \mathcal{B}_b(\mathcal{X})$, un sous-ensemble de $\mathcal{X}$ appartenant à la tribu $\mathcal{B}_b(\mathcal{X})$ restreinte aux ensembles bornés de $\mathcal{X}$. Un champ aléatoire Gamma $\Gamma(dx)$ est une mesure aléatoire non négative définie sur $\mathcal{X}$, caractérisée par une mesure de base $\alpha(dx)$ et un paramètre β telle que

1. $\Gamma(A) := \int_A \Gamma(dx) \sim \text{Gamma}(\alpha(A), \beta)$, avec $\alpha(A) := \int_A \alpha(dx)$;
2. pour tout $A_1, A_2 \in \mathcal{B}_b(\mathcal{X})$ tels que $A_1 \cap A_2 = \emptyset$, $\Gamma(A_1)$ et $\Gamma(A_2)$ sont des variables aléatoires indépendantes.

Concernant le noyau intervenant dans la construction $\Lambda(x) = \int K(x,x')\Gamma(dx')$, il peut être très général et des choix particuliers peuvent par exemple mener à des champs aléatoires non-stationnaires. Tous les choix de noyaux ne conviennent pas si l'on souhaite assurer des marginales



(a) (b)

FIGURE 2.6 – Objets spatio-temporels. À gauche : une ellipse $E(s, \gamma_1, \gamma_2, \phi)$ centrée en s. À droite : intersection entre 2 cylindres inclinés $A_{s,t}$ et $A_{s',t'}$ avec une durée δ.

Gamma. Pour répondre à cette dernière contrainte et limiter la complexité du modèle, nous avons fait le choix d'utiliser comme noyau une fonction indicatrice $K(x, x') = \mathbf{1}_A(x - x')$ où A est un cylindre elliptique incliné. L'ellipse centrée en $s = (s_1, s_2) \in \mathbb{R}^2$ est représentée avec ses paramètres sur la Figure 2.6-(a) et on note A_x le cylindre basé en $x = (s, t)$ (voir Figure 2.6-(b)). Dans la suite, nous considérons la mesure $\alpha(B) = \alpha\nu_d(B)/\nu_d(A)$, $B \in \mathcal{B}_b(\mathcal{X})$ avec $\nu_d(.)$ la mesure de Lebesgue sur $\mathbb{R}^d$. Par conséquent $\Lambda(x) \sim \text{Gamma}(\alpha, \beta)$, comme nécessaire pour le modèle (2.21).

Par ailleurs, la forme du noyau dans notre construction permet une interprétation directe de la structure de dépendance, et offre une interprétation physique du phénomène réel. L'un des intérêts d'utiliser ces objets est l'interprétation physique qui en découle. L'ellipse décrit la zone d'influence d'une tempête centrée en s. Ces ellipses (tempêtes) $E(s, \gamma_1, \gamma_2, \phi)$ se déplacent à travers l'espace avec une vitesse $\omega = (\omega_1, \omega_2) \in \mathbb{R}^2$ pour une durée $\delta > 0$ assurant une dynamique temporelle. De plus, nous obtenons des formules analytiques pour les distributions bivariées, ce qui facilite l'inférence statistique, l'interprétation et la caractérisation des propriétés de queue jointe.

2.4.3 Mesures de la dépendance extrémale associée au modèle

Comme le processus est stationnaire, il est possible de montrer que pour $(x, x') \in \mathcal{X}^2$, $x \neq x'$ et pour de grandes valeurs de v au-delà d'un seuil $u \geq 0$

$$\chi_{x,x'}(v) := \frac{\mathbb{P}(Z(x) > v, Z(x') > v)}{\mathbb{P}(Z(x') > v)} \sim 2^{-c_2} \left(\frac{v}{\beta}\right)^{c_2 - c_0}$$

avec $c_0 := \alpha(A_x)$ et $c_2 := \alpha(A_x \cap A_{x'})$. Comme $c_2 < c_0$, en faisant tendre v vers $+\infty$, il vient que $\chi_{x,x'} = 0$. Nous en déduisons que le processus Z est un processus asymptotiquement indépendant.

Pour caractériser la dépendance résiduelle, nous nous intéressons au coefficient $\bar{\chi}$. Nous obtenons

$$\bar{\chi}_{xx'}(v) := \frac{2\log \mathbb{P}(Z(x) > v)}{\log \mathbb{P}(Z(x) > v, Z(x') > v)} - 1 = \frac{2c_0}{c_1 + c_2 \dfrac{\log\left(1 + 2(v+k)/\beta\right)}{\log\left(1 + (v+k)/\beta\right)} + c_3} - 1$$

qui tend vers $\bar{\chi}_{x,x'} = \dfrac{c_2}{2c_0 - c_2}$, quand $v \to \infty$ avec c_0 et c_2 définis comme précédemment, $c_1 := \alpha(A_x \backslash A_{x'})$, $c_3 := \alpha(A_{x'} \backslash A_x)$ et donc $c_1 = c_3 = c_0 - c_2 \geq 0$.

Le coefficient $\bar{\chi}$ est donc égal au rapport entre le volume de l'intersection de A_x avec $A_{x'}$ et le volume de l'union de ces deux mêmes ensembles.

2.4.4 Mise en œuvre du modèle

Comme expliqué précédemment, la distribution bivariée associée au modèle est explicite dès lors que la transformée de Laplace bivariée du processus latent $\{\Lambda(x)\}$ l'est. Avec le processus choisi, le calcul de la distribution bivariée, elle-même basée sur le calcul d'intersections de deux cylindres rend tout à fait possible la mise en œuvre d'une inférence par vraisemblance par paires. Chaque paire d'observations possiblement censurées $Y(s_i, t), Y(s_j, t+k)$ contribue à la log-vraisemblance pondérée, le poids pouvant être nul. Pour plus de détails, il conviendra de se reporter à l'article lui-même (voir **[G3]**).

Le modèle a été mis en œuvre sur des données simulées et sur des données réelles. Dans les deux cas, l'accent a été mis sur l'étude de la dépendance spatio-temporelle. Aussi dans l'étude par simulations, les marges ont été supposées connues. Un plan de simulations a été proposé avec une configuration réaliste (30 sites et 2000 pas de temps). La méthode d'inférence par vraisemblance par paires donne de bons résultats d'estimation sur différents scénarios (avec ou sans vitesse, base circulaire ou elliptique plus ou moins prononcée).

Reprenant les données présentées en section 2.4.1, nous nous intéressons également à des données horaires de précipitation mesurées en 50 stations dans le sud de la France (voir Figure 2.4) entre 1993 et 2014 sur lesquelles nous n'étudions que les mois de septembre à novembre, ce qui représente au total 54542 heures.

En prenant pour seuil le quantile empirique à 99%, nous avons, et cela en chaque station, ajusté le modèle univarié (2.22). Nous sommes en présence d'une non-stationnarité spatiale évidente dans nos données. Notre objectif étant davantage centré sur la capacité du modèle proposé à capturer des dépendances spatio-temporelles, nous avons traité séparément la modélisation des marginales et celle de la structure de dépendance. Par conséquent, en utilisant les résultats de l'ajustement de la GPD en chaque site, les dépassements observés en chaque site ont été transformés pour satisfaire (2.22) avec $\xi = 1$ et $\sigma = \kappa + 1$.

Deux versions du modèle hiérarchique ont été ajustées, avec vitesse (G1) et sans vitesse (G2). Les résultats d'estimations se trouvent dans la Table 2.1. Le CLIC donne une préférence au modèle G1 mais la différence est assez peu marquée. Pour plus de détails notamment concernant la

comparaison avec d'autres modèles spatio-temporels asymptotiquement indépendants (processus gaussiens), nous renvoyons le lecteur à l'article **[G3]**.

	γ_1	γ_2	ϕ	δ	ω_1	ω_2
G1	165.062	318.823	0.085	20.184	0.723	0.446
	23.459	*19.811*	*0.026*	*0.948*	*0.195*	*0.009*
G2	175.817	294.323	0.041	20.036	0	0
	11.879	*25.291*	*0.064*	*1.039*	-	-

TABLE 2.1 – Estimations et écarts-types (en italique) des modèles ajustés. γ_1 et γ_2 sont donnés en kilomètres, ϕ en radians, δ en heures et ω_1 et ω_2 en kilomètres par heure.

Les durées estimées varient peu entre G1 et G2. Les estimations de l'angle ϕ diffèrent davantage mais cela est probablement lié au fait que les différences entre les semi-axes ne sont pas si marquées. De plus, les estimations de γ_1 et γ_2 sont similaires pour G1 et G2 et la vitesse est assez faible. Tous ces résultats présentent donc de la cohérence rendant possible leur interprétation.

L'appréciation de l'ajustement de ces modèles peut se faire à l'aide de diagnostiques graphiques. La Figure 2.7 montre les probabilités estimées $\mathbb{P}(Z(s,t) > v | Z(s',t') > v)$ dans différentes directions et pour différents écarts temporels $|t-t'|$. Les comportements des deux modèles sont effectivement très proches.

Pour aller plus loin, nous cherchons à comparer les distributions conditionnelles multivariées. Autrement dit, nous nous intéressons également à

$$\chi^*_{s_i;h}(v) := \mathbb{P}(Z(s_j,t) > v, s_j \in \partial s_i | Z(s_i, t-h) > v)$$

où ∂s_i est l'ensemble des 4 plus proches voisins du site s_i, $i = 1, \ldots, 50$.

Il s'agit de comparer les estimations empiriques $\hat{p}_i(h)$ de ces quantités avec celles sous le modèle ($\tilde{p}_i^{(j)}(h)$, $j = 1, \ldots, 200$ obtenues par Monte-Carlo et basées sur une procédure bootstrap sur 200 simulations de modèles G1 et G2). Pour chaque site s_i, l'erreur suivante (RMSE) est calculée

$$\mathrm{RMSE}_i(h) = \left\{ \frac{\sum_{j=1}^{200} (\tilde{p}_i^{(j)}(h) - \hat{p}_i(h))^2}{200} \right\}^{1/2},$$

ainsi qu'une mesure plus globale pour l'ensemble des sites $\mathrm{RMSE}(h) = \sum_{i=1}^{50} \mathrm{RMSE}_i(h)$. Les résultats pour $h = 0, 1, 2$ sont présentés en détail dans l'article pour les modèles G1 et G2 mais aussi pour un modèle alternatif construit sur la base d'un processus gaussien, et ce avec deux choix de q correspondant aux quantiles à 99% et 99.5%. En considérant le seuil le plus élevé ($q_{0.995}$), les modèles G1 et G2 présentent de meilleurs ajustements que le modèle alternatif.

2.5 Perspectives

Dans ce chapitre dédié à la prise en compte de l'indépendance asymptotique, j'ai présenté des modèles stochastiques pour les extrêmes spatiaux et spatio-temporels et leur analyse théorique

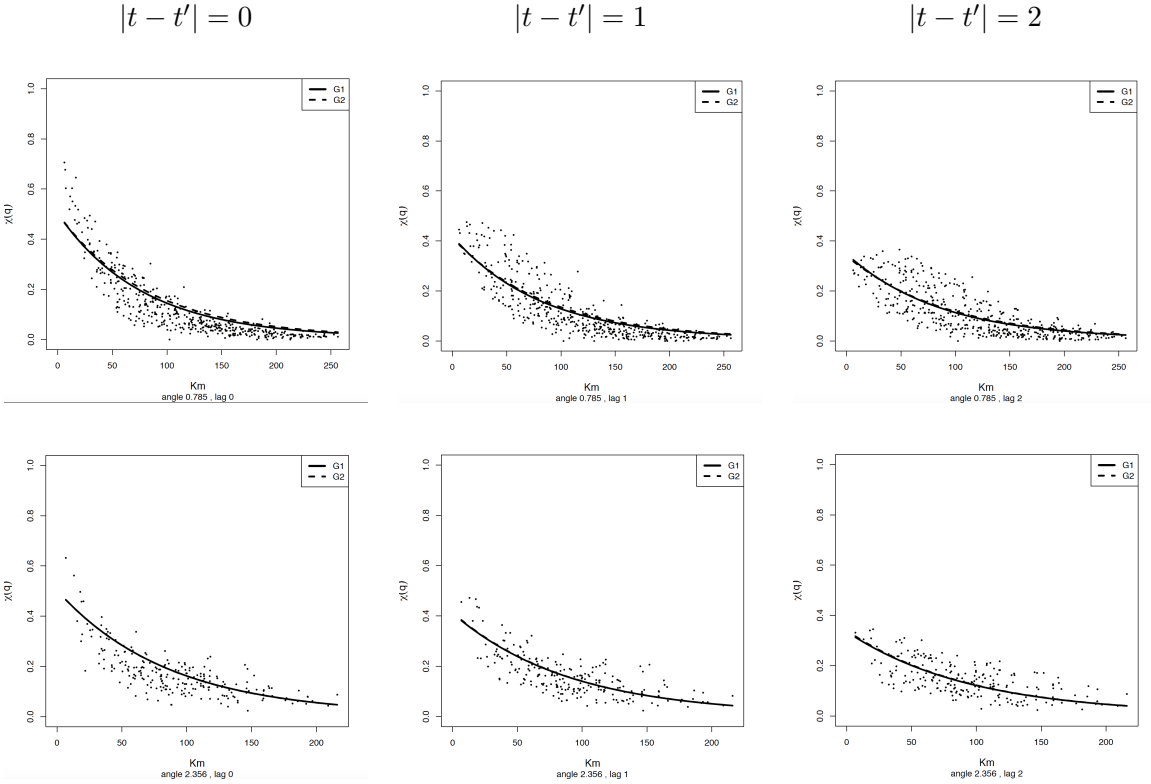


FIGURE 2.7 – Probabilités estimées $\mathbb{P}(Z(s,t) > v | Z(s',t') > v)$ en fonction de la distance $||s-s'||$. Chaque ligne correspond à une direction différente et chaque colonne à un écart temporel $|t-t'|$ différent. Les points correspondent aux estimations empiriques. La valeur v correspond ici au quantile à 99%.

que ce soit en terme d'interprétation ou de propriétés associées. Je propose également l'utilisation d'outils statistiques permettant notamment l'inférence des paramètres des modèles. Ces étapes sont indispensables et permettent ensuite de mettre ces modèles en application sur des données de pluies. Je m'intéresse à l'ensemble de ces étapes essayant de prendre en compte toute la complexité inhérente aux données spatiales ou spatio-temporelles extrêmes.

Le premier modèle proposé est un modèle spatial qui, combinant une composante asymptotiquement indépendante et une composante asymptotiquement dépendante, présente une grande flexibilité du type de dépendance présenté. En effet, le processus construit peut ainsi avoir toutes ses paires asymptotiquement indépendantes ou bien toutes ses paires asymptotiquement dépendantes ou encore présenter les deux types de dépendance et ce, en fonction de la distance entre les sites. Présenter des types de dépendance différents en fonction de la distance entre les sites constitue un objectif qui motive des travaux très récents voire actuels (Tawn *et al.* , 2018; Wadsworth & Tawn, 2019) basés sur les approches conditionnelles (Heffernan & Tawn, 2004). D'autres propositions de modèles spatiaux convenant à la fois pour des cas asymptotiquement dépendants ou asymptotiquement indépendants ont été récemment faites (Huser *et al.* , 2017; Huser & Wadsworth, 2019) mais nécessitent d'avoir le même type de dépendance partout dans la zone considérée.

Le second modèle que nous avons proposé est un modèle hiérarchique pour les dépassements

possédant de nombreux avantages. Il s'agit d'un modèle spatio-temporel anisotrope. Les paramètres de la structure de dépendance ont une interprétation physique et peuvent être estimés par vraisemblance composite. De plus il est asymptotiquement indépendant dans le temps et dans l'espace.

Les deux modèles proposés et détaillés dans ce chapitre s'inspiraient ou s'appuyaient sur d'autres modèles et serviront eux-mêmes sans aucun doute de base à la conception de nouvelles modélisations. Parmi elles, une piste que nous souhaitons poursuivre avec J.N. Bacro, J. Carreau et C. Gaetan concerne l'extension du modèle de max-mélange. Le paramètre de mélange β pourrait dépendre de l'espace, à l'instar de a dans la proposition présentée dans **[G1]** et présentée en annexe D de ce document. Cela permettrait ainsi une non-stationnarité spatiale de la structure de dépendance. Il est en effet tout à fait pertinent d'imaginer que le mélange soit différent selon les zones géographiques, notamment si l'on travaille sur des zones élargies au relief hétérogène. Cette perspective est reprise et détaillée au chapitre 4 section 4.1. Nous pourrions également imaginer un modèle non-stationnaire dans le temps. Cela nécessiterait bien sûr de proposer une extension spatio-temporelle du modèle de max-mélange avec un paramètre de mélange dépendant également de l'espace et du temps. Les objectifs poursuivis seraient alors multiples. Outre le fait de prendre en compte une évidente dépendance temporelle, nous pourrions imaginer une modélisation permettant des types de dépendance différentes dans le temps et dans l'espace.

Concernant le second modèle présenté qui est lui spatio-temporel, des réflexions pour être capable de considérer la dépendance asymptotique dans le modèle ont été menées. La dépendance asymptotique dans notre construction (2.21) est équivalente à présenter une dépendance dans la queue inférieure de $\Lambda(x)$. Une idée prometteuse pour introduire cela est la suivante : étant donné un processus temporel $B(t)$ indépendant de $\Lambda(s,t)$ avec des marginales $Beta(\widetilde{\alpha},\alpha)$, $0 < \widetilde{\alpha} < \alpha$, on peut remplacer dans la construction $\Lambda(s,t)$ par $\widetilde{\Lambda}(s,t) = B(t)\Lambda(s,t)$ de marginales $\Gamma(\widetilde{\alpha},\beta)$. Cette construction rend le modèle asymptotiquement dépendant dans l'espace et il pourrait l'être également dans le temps en fonction de la dépendance dans la queue inférieure de $B(t)$.

Chapitre 3

Approches semi-paramétriques pour la simulation d'événements spatio-temporels extrêmes

Sommaire

3.1 Introduction

Être en capacité de générer des épisodes spatio-temporels extrêmes de précipitations ou de vagues, dont on peut quantifier la force, permet notamment d'étudier l'impact de ces "scénarios catastrophes". Réalisant le déficit de méthodes permettant des simulations réalistes d'extrêmes, c'est-à-dire dont la simulation peut réellement se ramener à un événement observable, j'ai fait de cette problématique une de mes orientations de recherche privilégiée. Au cours de la thèse de Chailan (2015), nous appuyant sur le travail de Ferreira & de Haan (2014), nous avons

notamment proposé une approche semi-paramétrique pour la simulation de processus spatio-temporels de vagues **[G6]** (voir section 3.2). Plus précisément, et même si nous utiliserons tout au long de cette partie le terme *simulation*, il s'agira de méthodes d'*amplification*. En effet, la méthode est totalement non-paramétrique pour la structure de dépendance, l'idée étant de s'appuyer sur les motifs spatiaux présents dans les épisodes extrêmes extraits des données. Cette approche permet, pour un épisode extrême extrait des données, de produire de nouveaux épisodes à des niveaux plus extrêmes. Mais pour un niveau extrême attendu (correspondant par exemple à une certaine période de retour en un site), à un épisode extrait ne pourra correspondre qu'un seul nouvel épisode. C'est en ce sens que les simulations produites par la méthodologie proposée dans **[G6]** sont en nombre limité. J'ai cherché d'une part à lever ce verrou et d'autre part à rapprocher cette première proposition du formalisme des processus de Pareto. Au-delà de la complexité évidente engendrée par l'aspect à la fois spatial et temporel des approches proposées, une difficulté de travailler en spatio-temporel réside dans la définition même d'un épisode spatio-temporel extrême. Ce dernier peut être associé à une valeur très forte de la variable d'intérêt en un point de l'espace et en un temps donné ou encore à des valeurs plus faibles mais dont le cumul sur une période de temps donné et sur une zone est élevé. Pour préciser la notion d'événement extrême spatio-temporel, nous nous intéressons aux dépassements d'une fonctionnelle de coût homogène notée ℓ. Plus précisément, nous proposons de nous appuyer sur le formalisme des processus de Pareto (Ferreira & de Haan, 2014; Dombry & Ribatet, 2015; De Fondeville & Davison, 2018) que nous réécrivons dans un cadre spatio-temporel (voir section 3.3), ces derniers étant parfaitement adaptés pour modéliser des phénomènes dépassant un seuil suffisamment élevé. Une nouvelle proposition pour simuler des épisodes extrêmes spatio-temporels **[G15, G20]**, améliorant la première expliquée en section 3.2 (voir aussi **[G6]**), est formulée et est en partie décrite en section 3.4. L'approche proposée reste non-paramétrique pour la structure de dépendance et garde donc l'intérêt d'être complètement guidée par les données sans modèle a priori. De plus elle reste facilement paramétrable et permet donc de contrôler le niveau extrême de l'épisode simulé aléatoirement.

Les principaux travaux scientifiques présentés dans ce chapitre ont été faits dans le cadre du co-encadrement d'une thèse et d'un post-doctorat (respectivement R. Chailan 2012-2015 et F. Palacios-Rodriguez 2017-2019). Ma contribution personnelle va néanmoins au-delà de la simple direction de ces travaux et mon investissement a été conséquent, le point de départ étant tout naturellement les propositions précises des objectifs et moyens à mettre en œuvre pour les atteindre. Par ailleurs, la proposition, l'animation et la gestion des projets associés à ces thèmes est une contribution personnelle au service du collectif ayant permis la création d'un groupe de recherche fédéré autour de questions scientifiques communes et organisées pour répondre à un réel défi et besoin : simuler des processus environnementaux spatio-temporels réalistes et intégrant des extrêmes pour, in fine, mesurer leurs impacts. Sur ces thématiques, j'ai en effet pu obtenir le financement de deux projets LEFE rassemblant pour le second 14 chercheurs répartis sur 9 partenaires. Ces projets nous ont apporté de la visibilité et des moyens et ont favorisé les échanges, collaborations et diffusion de résultats. Cette contribution touche des communautés différentes. D'une part, il s'agit bien sûr de travaux de recherche originaux en statistique (1 article publié à Annals of Applied Statistics **[G6]**, un article soumis en 2019 **[G20]**, un chapitre de livre **[G14]** et 3 proceedings **[G16, G18, G19]**). Les articles **[G6, G20]** se trouvent en annexe de ce document (partie III annexes B et E). Par ailleurs, dans **[G6]**, cette méthodologie innovante appliquée sur des données de vagues a été mise au service de l'étude du risque côtier alors que dans **[G20]** nous nous sommes intéressés aux épisodes de pluie dans la région de Montpellier. Cela a permis des rapprochements entre d'une part l'Institut Montpellierrain Alexander Grothendieck (IMAG)

et d'autre part Géosciences-M (participation au réseau GLADYS) et HydroSciences Montpellier (participation de FRAISE à l'observatoire urbain de Montpellier). J'ai porté la rédaction d'un chapitre de livre paru en 2020 **[G13]** sur les simulations de pluies et les inondations urbaines. Cette thématique de recherche est au centre de mon projet de recherche et de fait de celui de l'équipe projet Inria LEMON dont je suis membre depuis le 1er septembre 2017 et qui a été officiellement créée en janvier 2019.

3.2 Première proposition de méthode effective pour la simulation d'événements extrêmes spatio-temporels.

3.2.1 Contexte

Une première contribution sur la simulation d'événements spatio-temporels extrêmes a été menée dans le cadre de la thèse de Chailan (2015) que j'ai co-encadrée. Le travail réalisé dans ce projet se situait à l'interface de l'analyse statistique, de la géophysique et de l'informatique et avait pour objectif d'apporter des méthodologies et outils aux décideurs en charge de la gestion de risques côtiers. Cette thèse s'est effectuée dans le cadre d'une collaboration entre l'IMAG, le Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), Géosciences Montpellier et IBM Montpellier et portait sur l'application du calcul scientifique et de l'analyse statistique à la gestion du risque inondation en milieu littoral. Les risques littoraux sont généralement des conséquences de conditions environnementales extrêmes qui sont rarement observées. L'énergie véhiculée par les vagues est la principale responsable des risques littoraux comme l'érosion et la submersion. Le nombre de bouées en mer mesurant ces hauteurs de vagues est très limité (4 dans le golfe du Lion). C'est pourquoi nous avons mis en œuvre une simulation numérique (basée sur le modèle WaveWatch 3) nous dotant ainsi d'un jeu de données de hauteurs significatives de vagues à haute résolution spatiale et temporelle dans le golfe du Lion. Ces données, y compris les plus extrêmes, ont été validées grâce aux mesures réalisées aux quatre bouées disponibles sur la zone. Une première étude a mis en évidence la nécessité de prendre en compte la dépendance extrémale spatiale et temporelle présente dans ces données **[G18]**. Par ailleurs, sur la base de ce jeu de données et nous appuyant notamment sur les travaux de Caires *et al.* (2011); Groeneweg *et al.* (2012); Ferreira & de Haan (2014), nous avons proposé une méthode semi-paramétrique dont le but est de construire de nouveaux épisodes qui soient plus extrêmes encore que ceux observés. Certains éléments de cette méthode sont donnés en section 3.2.2 et pour plus de détails, nous renvoyons le lecteur à l'article **[G6]** également disponible en annexe B de ce document. Les épisodes extrêmes créés par cette méthode peuvent alors être vus comme des forçages d'états-de-mer extrêmes permettant ensuite d'alimenter des modèles numériques à la côte pour étudier les risques associés. Enfin, l'idée consistant à pré-calculer des scénarios extrêmes et leurs risques associés puis à les stocker constitue une piste prometteuse dans une perspective d'aide à la décision en temps réel et illustre l'intérêt d'associer des méthodes numériques, statistiques et informatiques (voir **[G19]**).

3.2.2 Descriptif, originalité et force de la méthode

Ayant pour objectif de proposer une méthode de simulation d'événements spatio-temporels extrêmes, nous décrivons une méthodologie basée sur une technique d'amplification de tempêtes réelles (observées). Il s'agit d'une approche semi-paramétrique qui s'appuie fortement sur Caires *et al.* (2011); Groeneweg *et al.* (2012); Ferreira & de Haan (2014). Aucun modèle n'est utilisé pour la structure de dépendance contrairement à ce qui est fait pour les marges. Cette approche est basée sur les dépassements d'un seuil élevé et diffère donc des approches basées sur les maximas. Ces dernières reposent sur l'étude des processus max-stables dont les simulations stochastiques sont possibles (voir notamment les travaux de Dombry *et al.* , 2013, 2016; Oesting *et al.* , 2018; Oesting & Stein, 2018). Néanmoins, l'interprétation physique de réalisations de processus max-stables est difficile car elles agrègent l'information de plusieurs événements sous-jacents.

La méthode proposée se décompose en 4 étapes. La première étape est une étape de prétraitement et consiste à ramener l'ensemble des données à une échelle Pareto standard via une transformation notée T. La seconde étape consiste, à partir des données ainsi standardisées, à extraire ce que l'on qualifie de *tempêtes*. Les tempêtes sont les épisodes extrêmes extraits à partir des données standardisées. Dans ce premier travail, il s'agit du processus standardisé en tout point de l'espace et considéré sur une fenêtre temporelle réduite. Une fois les tempêtes extraites, elles sont amplifiées dans une troisième étape vers des valeurs plus élevées de manière contrôlée à l'aide d'un coefficient $\zeta > 1$. Enfin la quatrième et dernière étape consiste à ramener les tempêtes amplifiées à leur échelle d'origine au moyen de la fonction inverse de T.

L'approche est asymptotiquement justifiée (voir Caires *et al.* (2011); Groeneweg *et al.* (2012); Ferreira & de Haan (2014) et **[G6]**). Une condition à l'utilisation de cette approche est que le processus d'intérêt soit dans le domaine d'attraction max-stable ce qui suppose donc d'être en situation de dépendance asymptotique. L'idée de la méthode est de réutiliser la structure de dépendance présente dans les tempêtes extraites. Cela est tout à fait pertinent car sous l'hypothèse de dépendance asymptotique, la dépendance reste stable pour tous les niveaux extrêmes considérés. Autrement dit, si l'on sélectionne une tempête suffisamment extrême, il s'agit d'une réalisation de la structure de dépendance extrémale qui peut, de fait, être réutilisée à des niveaux plus extrêmes.

Notre objectif est de produire des états-de-mer extrêmes. Autrement dit, nous nous intéressons à un processus spatio-temporel bivarié composé de deux quantités d'intérêt, la hauteur significative de vagues et la période. Pour vérifier le contexte de dépendance asymptotique, nous avons estimé un coefficient extrémal dans différents contextes. Plus précisément nous avons d'une part estimé la fonction coefficient extrémal selon la distance entre les paires des sites. Il en résulte que l'on peut, et ce quelle que soit la distance, donc sur toute la zone, supposer raisonnablement être en situation de dépendance asymptotique. D'autre part, nous l'avons également considéré en fonction d'un écart temporel pour nous assurer de la dépendance asymptotique dans le temps et surtout de la durée maximale des épisodes pour rester sous cette hypothèse. Il apparait que l'hypothèse de dépendance asymptotique est raisonnable pour des tempêtes d'une durée d'environ 50 heures maximum. Enfin, nous nous sommes assurés de la dépendance extrémale entre la hauteur significative des vagues et la période quelle que soit la bathymétrie.

L'approche proposée dans **[G6]** a été utilisée pour simuler des processus extrêmes spatio-

temporels bivariés (hauteur des vagues, période), dans le but d'étudier les conséquences d'états-de-mer extrêmes. Notre approche se démarque de ce qui a été proposé dans la littérature notamment par Caires *et al.* (2011); Groeneweg *et al.* (2012); Ferreira & de Haan (2014) sur certains aspects comme le cadre bivarié ou encore la sélection des tempêtes. Ces dernières sont sélectionnées selon l'un des deux processus, les hauteurs de vagues dans notre cas. En effet, le processus le plus extrême est défini comme celui ayant à un instant donné la hauteur de vagues (standardisée) la plus élevée sur la zone littorale (définie comme un sous-ensemble de la zone d'étude). La sélection des hauteurs de vagues et périodes sur toute la zone et pendant 24h (le pic étant supposé au milieu de l'épisode) constitue l'épisode bivarié spatio-temporel le plus extrême. En d'autres termes il s'agit de la première tempête sélectionnée (i.e., la plus extrême au sens de la hauteur des vagues en un site). A partir de cette dernière, d'autres épisodes plus extrêmes encore, pourront être créés par *uplift* au moyen d'un vecteur de paramètres (un pour les hauteurs et un pour les périodes) garantissant un certain niveau d'amplification comme par exemple une certaine période de retour en le site où la première tempête a été sélectionnée. De la même manière, les m tempêtes les plus extrêmes - tant qu'elles peuvent être définies comme telles - pourront être sélectionnées pour servir de base à la construction d'autres épisodes.

Comme cela est stipulé précédemment, cette méthode de simulation est directement inspirée de l'approche constructive des processus de Pareto (Ferreira & de Haan, 2014; Dombry & Ribatet, 2015). Comme détaillé dans les sections suivantes, nous avons proposé d'aller plus loin dans ce rapprochement pour proposer une méthode d'amplification plus générale, permettant notamment à partir d'une tempête extraite de simuler une infinité d'épisodes extrêmes.

3.3 Processus de Pareto spatio-temporels

Soit $\mathcal{S}$ un sous-ensemble compact de $\mathbb{R}^d$ qui représentera une zone spatiale d'intérêt (en pratique, $d = 2$) et $\mathcal{T}$ un sous-ensemble compact de $\mathbb{R}^+$ pour la dimension temporelle. Nous notons $C(\mathcal{S} \times \mathcal{T})$ l'espace des fonctions continues sur $\mathcal{S} \times \mathcal{T}$. La restriction de $C(\mathcal{S} \times \mathcal{T})$ aux fonctions non-négatives s'écrit $C_+(\mathcal{S} \times \mathcal{T})$. De la même manière, on définit l'espace des fonctions non-négatives continues dans $\mathcal{S}$ par $C_+(\mathcal{S})$.

En théorie des valeurs extrêmes multivariée, la loi de Pareto généralisée a été introduite par Rootzén & Tajvidi (2006) et apparait comme loi limite en conditionnant par le fait qu'au moins une des composantes dépasse un seuil élevé. Cette idée a été étendue au cas infini-dimensionnel par Ferreira & de Haan (2014) qui introduisent les processus de Pareto pour lesquels la condition est basée sur des dépassements du supremum sur le domaine d'étude. Pour gagner en flexibilité, Dombry & Ribatet (2015) et De Fondeville & Davison (2018) considèrent la notion de processus ℓ-Pareto en considérant des dépassements plus généraux définis pour une fonctionnelle de coût homogène notée ℓ. En suivant De Fondeville & Davison (2018), on pourra également parler de fonctionnelle de risque (notée alors r par les auteurs) puisqu'elle détermine le type d'événements extrêmes dont on souhaite étudier le risque.

Nous nous concentrons ici sur les dimensions spatiales et temporelles de l'étendue des événements extrêmes. Puisque nous visons à modéliser des phénomènes qui dépassent un certain seuil extrême, nous commençons par définir et caractériser les processus ℓ-Pareto dans un cadre spatio-temporel. La définition constructive suivante s'appuie sur Dombry & Ribatet (2015).

3.3.1 Construction de processus ℓ-Pareto spatio-temporels

Soit une fonctionnelle de coût $\ell : C_+(\mathcal{S} \times \mathcal{T}) \to [0, +\infty)$ une fonction non négative continue qui est homogène, i.e., $\ell(tf) = t\ell(f)$ pour $t \geq 0$. Des exemples possibles pour ℓ sont les fonctions suivantes : maximum, minimum, moyenne ou encore la valeur en un point spécifique $(s_0, t_0) \in \mathcal{S} \times \mathcal{T}$.

Définition 3.3.1 (Processus ℓ-Pareto spatio-temporel standard)
Soit $W^* = \{W^*(s,t)\}_{s\in\mathcal{S},t\in\mathcal{T}}$ un processus stochastique dans $C_+(\mathcal{S} \times \mathcal{T})$. On appelle W^* un processus ℓ-Pareto spatio-temporel standard s'il peut être représenté de la manière suivante

$$W^*(s,t) \stackrel{d}{=} RY(s,t) \tag{3.1}$$

où

1. *Y est un processus stochastique dans $C_+(\mathcal{S} \times \mathcal{T})$ satisfaisant $\ell(Y) = 1$;*
2. *R est une variable aléatoire distribuée selon une loi de Pareto avec un paramètre d'échelle égal à 1 et un paramètre de forme γ_R, i.e., $\mathbb{P}(R > r) = r^{-\gamma_R}$, $r > 1$;*
3. *Y et R sont indépendants.*

La définition précédente est équivalente à la définition par la propriété de stabilité POT : pour tout $u \geq 1$, la distribution du processus renormalisé $\{u^{-1}W^*|\ell(W^*) \geq u\}$ est égale à la distribution de W^* ; voir Théorème 2 dans Dombry & Ribatet (2015). Par construction, on obtient $Y \stackrel{d}{=} W^*/\ell(W^*)$ et $R \stackrel{d}{=} \ell(W^*)$. Une version généralisée des processus de ℓ-Pareto spatio-temporels est donnée en définition 3.3.2 pour apporter de la flexibilité au niveau des marginales.

Définition 3.3.2 (Processus ℓ-Pareto spatio-temporel généralisé)
En considérant un processus ℓ-Pareto $W^*(s,t)$ construit suivant la Définition 3.3.1 et des fonctions réelles et continues $\sigma(s,t) > 0$, $\mu(s,t)$ et $\gamma(s,t)$ dans $C(\mathcal{S} \times \mathcal{T})$, un processus ℓ-Pareto spatio-temporel généralisé est un processus construit de la façon suivante

$$W(s,t) \stackrel{d}{=} \begin{cases} \mu(s,t) + \sigma(s,t)\{W^*(s,t)^{\gamma(s,t)} - 1\}/\gamma(s,t), & \gamma(s,t) \neq 0, \\ \mu(s,t) + \sigma(s,t)\log W^*(s,t), & \gamma(s,t) = 0. \end{cases} \tag{3.2}$$

3.3.2 Résultats asymptotiques pour les processus ℓ-Pareto spatio-temporels

Nous rappelons ci-après les deux principaux résultats asymptotiques pour caractériser les extrêmes de processus stochastiques : les processus max-stables et les processus de Pareto. Pour les détails techniques, nous renvoyons le lecteur à la littérature (Lin & de Haan, 2001; de Haan & Ferreira, 2006; Ferreira & de Haan, 2014; Thibaud & Opitz, 2015; Dombry & Ribatet, 2015; De Fondeville & Davison, 2018).

Dans la suite, nous utilisons le symbole "$\Rightarrow$" pour différentes variantes de la convergence faible d'éléments aléatoires du domaine univarié, multivarié ou spatial.

Soit $X_1, \ldots, X_n$, des copies indépendantes d'un processus stochastique spatio-temporel $X = \{X(s,t)\}_{s\in\mathcal{S},t\in\mathcal{T}}$ avec des trajectoires continues. On dit que le processus X est dans le domaine d'attraction d'un processus max-stable $Z = \{Z(s,t)\}_{s\in\mathcal{S},t\in\mathcal{T}}$ avec des trajectoires continues si il existe des fonctions continues $a_n > 0$ et b_n telles que

$$\left\{\max_{1\le i\le n} \frac{X_i(s,t) - b_n(s,t)}{a_n(s,t)}\right\}_{s\in\mathcal{S},t\in\mathcal{T}} \Rightarrow \{Z(s,t)\}_{s\in\mathcal{S},t\in\mathcal{T}}\,. \tag{3.3}$$

Plus de détails sur les processus max-stables spatio-temporels se trouvent dans les travaux de Davis *et al.* (2013a,b).

Les convergences de la structure de dépendance et des distributions marginales dans (3.3) peuvent être étudiées séparément ; voir de Haan & Ferreira (2006, Section 9.2). Un processus standardisé $X^* = \{X^*(s,t)\}$ peut être défini comme $X^*(s,t) = H^{-1}(F_{(s,t)}(X(s,t)))$, $s \in \mathcal{S}$, $t \in \mathcal{T}$, où H^{-1} correspond à la fonction inverse de la fonction de répartition de la Pareto standard H, et $F_{(s,t)}$ est la fonction de répartition de $X(s,t)$. Si X a des distributions marginales continues $F_{(s,t)}$, alors X^* a des marginales Pareto standard. Pour $a_n \equiv n,\ b_n \equiv 0$, la limite max-stable pour X^* dans (3.3) est un processus max-stable de marginales Fréchet standard (processus max-stable standard) $Z^* = \{Z^*(s,t)\}_{s\in\mathcal{S},t\in\mathcal{T}}$; voir de Haan & Ferreira (2006, Définition 9.2.4).

Si X^* est dans le domaine d'attraction d'un processus max-stable Z^*, on obtient la convergence des ℓ-*excès* sur l'échelle standard :

$$\left\{u^{-1}X^*(s,t) | \ell(X^*(s,t)) > u\right\} \Rightarrow \{W^*(s,t)\}\,,\ u \to \infty, \tag{3.4}$$

où $W^*(s,t)$ est un processus ℓ-Pareto spatio-temporel standard comme dans la Définition 3.3.1 (Dombry & Ribatet, 2015, Théorème 3). Inversement, si la convergence dans (3.4) est vérifiée pour ℓ correspondant au maximum, alors nous avons la convergence dans (3.3) du processus max-stable X^* vers Z^*.

3.3.3 Discussions de quelques points concernant la mise en pratique des processus de Pareto

En pratique, les résultats asymptotiques présentés précédemment sont utilisés dans le cadre d'échantillons de taille finie. Certains choix doivent alors être faits et nous faisons, dans cette section, des propositions concernant trois problématiques. Des détails et compléments se trouvent dans l'article **[G20]** soumis pour publication en 2019, dont une version est présentée en annexe E de ce document.

Transformations marginales

Nous commençons par discuter les transformations marginales de X telles que X^* satisfasse (3.4). Les difficultés se posent souvent en pratique pour les petites valeurs. Par exemple, les faibles valeurs comme la valeur 0, qui arrivent avec une probabilité non négligeable dans notre étude correspondant alors à l'absence de pluie, doivent pouvoir correspondre également à un 0 pour le processus standardisé X^* (pour l'application, voir section 3.5).

D'une manière générale, nous faisons le choix de la distribution $G : \mathbb{R} \to [0, 1]$ dont la fonction de survie $\bar{G}$ vérifie : $x\,\bar{G}(x) \to 1,\ x \to \infty$, et $\bar{G}(0) = 1$; on note $G^{\leftarrow}$ pour l'inverse (généralisé) de G. On définit alors la transformation $T = T_{(s,t)} : \mathbb{R} \to [0, \infty)$ vers le processus standardisé X^* comme suit :

$$X^*(s,t) = T(X(s,t)) = G^{\leftarrow}(F_{(s,t)}(X(s,t))) \tag{3.5}$$

où $F_{(s,t)} : \mathbb{R} \to [0, 1]$ est la distribution of $X(s, t)$. La transformation inverse de T peut alors se définir comme $T^{\leftarrow}(f) = F^{\leftarrow}_{(s,t)}(G(f))$ pour $f \in C_+(\mathcal{S} \times \mathcal{T})$, avec $F^{\leftarrow}_{(s,t)}$ l'inverse (généralisé) de $F_{(s,t)}$.

Concernant le choix de $F_{(s,t)}$, il est naturel d'utiliser un modèle de queue motivé par la théorie univariée des valeurs extrêmes dont la paramétrisation correspond directement à celle du processus de Pareto dans la définition 3.3.2. Pour un seuil $u(s, t)$ élevé on suppose que

$$\mathbb{P}(X(s,t) > x) = 1 - F_{(s,t)}(x) = \left[1 + \gamma(s,t)\frac{x - \mu(s,t)}{\sigma(s,t)}\right]_+^{-1/\gamma(s,t)} \tag{3.6}$$

pour $x > u(s, t)$, avec les fonctions paramètre de localisation $\mu(s, t) < u(s, t)$, d'échelle $\sigma(s, t) > 0$ et de forme $\gamma(s, t)$, et tel que la partie droite de (3.6) est inférieure à 1 (Thibaud & Opitz, 2015). Pour les valeurs $X(s, t)$ inférieures à $u(s, t)$, on peut utiliser la fonction de répartition empirique ou d'autres distributions satisfaisant le fait que la probabilité d'être inférieure à $u(s, t)$ correspond à $F_{(s,t)}(u(s, t))$ avec $F_{(s,t)}$ définie en (3.6).

En utilisant la standardisation dans (3.5), il vient que $\mathbb{P}(T(X(s, t)) > T(x)) \sim \frac{1}{T(x)}$ pour x grand, et par conséquent nous avons $\mathbb{P}(T(X'(s, t)) > T(X(s, t)) \mid X(s, t) = x(s, t)) \sim \frac{1}{T(x(s,t))}$ pour une copie indépendante X' de X. Pour $X(s, t)$ observé, la valeur de $T(X(s, t))$ peut être interprétée comme la période de retour (marginale) de l'observation $X(s, t)$, et pour des quantiles élevés, on peut interpréter X^* comme le processus spatio-temporel des périodes de retour marginales.

Définition d'événements spatio-temporels extrêmes

Si notre objectif est de simuler des scénarios extrêmes spatio-temporels, il est impératif de définir ce qu'on entend alors par *extrême*. Il n'existe pas de définition unique de ce qu'est un événement extrême, c'est-à-dire de définition de la fonctionnelle de coût ℓ. Cette dernière dépend plutôt de la nature du phénomène considéré, de l'ensemble de données ou encore de l'objectif de l'étude. Les connaissances d'experts peuvent suggérer comment mesurer la nature extrême d'un événement, où la question de savoir comment combiner les critères liés à la durée, l'étendue spatiale et l'ampleur est récurrente. Dans un cadre spatial et paramétrique, l'application sur les pluies en Floride proposée par De Fondeville & Davison (2018) et effectuée avec différentes fonctionnelles permet de considérer différents types de pluies (locales et intenses ou accumulation de pluies plus étalées dans l'espace). Les auteurs soulèvent également le problème de travailler sur des données transformées et les soucis d'interprétabilité que cela peut générer. Avec les données environnementales, nous n'avons souvent qu'une seule observation du processus spatio-temporel X, et les valeurs très élevées ont généralement tendance à se regrouper dans le temps et à former des clusters, des sous-périodes relativement courtes. Nous considérons ces sous-périodes comme des événements (ou épisodes) spatio-temporels extrêmes dont la force est quantifiée au moyen de

la fonctionnelle de coût. En pratique, nous appliquons cette fonctionnelle à une large collection d'épisodes candidats pour en extraire les plus extrêmes. Notre algorithme d'extraction, détaillé en section 3.4.1 est conçu pour éviter l'intersection temporelle des épisodes extrêmes sélectionnés.

Nous utilisons l'idée de fenêtres spatio-temporelles glissantes et spécifions le support de la fonction de coût ℓ introduite dans la section 3.3.1 comme un voisinage $\mathcal{N}(s,t)$ de la position $s \in \mathcal{S}$ et du temps $t \in \mathcal{T}$. En pratique, la taille de la fenêtre définit la durée maximale et l'étendue spatiale des événements extrêmes. Ce voisinage pourrait être défini par une durée δ dans le temps, et le support spatial pourrait être la zone d'étude complète (dans ce cas on omettrait l'indice spatial s) ou une sous-région telle que la zone littorale ou définie par une certaine distance autour d'un site spécifique s_0. Pour indiquer le support local de la fonction de coût définie sur un voisinage de (s,t), nous utilisons la notation $\ell_{s,t}(X^*) = \ell(\{X^*(s',t'),\ (s',t') \in \mathcal{N}(s,t)\})$.

On propose de définir $\mathcal{N}(s,t)$ comme le produit d'un voisinage spatial $\mathcal{N}(s)$ et d'un voisinage temporel $\mathcal{N}(t)$ (comme $\{t' \in \mathcal{T} \mid |t-t'| \leq \delta \text{ heures}\}$), $\mathcal{N}(s,t) = \mathcal{N}(s) \times \mathcal{N}(t)$. Des fonctionnelles de coût utiles ℓ pour des épisodes spatio-temporels sont obtenues par composition d'une fonctionnelle spatiale ℓ^S avec une fonctionnelle temporelle ℓ^T, cette dernière s'appliquant aux valeurs de ℓ^S observées sur δ pas de temps successifs :

$$\ell_{s,t}(X^*) = \ell^T(\ell^S_{s,t-(\delta-1)}(X^*), \ldots, \ell^S_{s,t}(X^*)), \tag{3.7}$$

avec $\ell^S_{s,t}(X^*) = \ell^S(\{X^*(s',t) \mid s' \in \mathcal{N}(s)\})$ et δ la durée de l'épisode.

Si X^* satisfait la condition d'appartenance au domaine d'attraction fonctionnelle (3.3), alors

$$\mathbb{P}(\ell(X^*) > u) \sim \theta_\ell / u, \quad u \to \infty, \tag{3.8}$$

où θ_ℓ est le *coefficient ℓ-extrémal* (voir Engelke *et al.* , 2019). Quand $\ell_{s,t}$ correspond au maximum spatio-temporel sur $\mathcal{N}(s,t)$ (i.e., $\ell^T = \max$ et $\ell^S_{s,t} = \max$), le coefficient ℓ-extrémal $\theta_{\ell_{s,t}}$ correspond au coefficient extrémal classique sur le domaine $\mathcal{N}(s,t)$ (voir Exemple 4 de Engelke *et al.* , 2019). Avec $(s_0,t_0) \in \mathcal{S} \times \mathcal{T}$ un point spatio-temporel fixé, si l'on définit la fonctionnelle de coût $\ell(X^*)$ comme $X^*(s_0,t_0)$, alors on a $\theta_\ell = 1$. Par ailleurs, θ_ℓ est également égal à 1 si ℓ correspond à la moyenne, c'est-à-dire si $\ell_{s,t}(x) = \frac{1}{|\mathcal{N}(s,t)|} \int_{\mathcal{N}(s,t)} x(s',t') \, \mathrm{d}(s',t')$; voir Ferreira *et al.* (2012, Proposition 2.2).

Comme expliqué précédemment, il est possible d'interpréter X^* comme le processus spatio-temporel des périodes de retour marginales. La fonctionnelle de coût ℓ vient alors agréger ces périodes de retour marginales $X^*(s,t)$.

Par ailleurs, en utilisant (3.8), on peut approximativement calculer des niveaux de retour pour des épisodes extrêmes caractérisés comme des excès de ℓ par rapport à un seuil élevé u. Comme $\mathbb{P}(\ell((X')^*) > \ell(X^*) \mid X^* = x^*) \sim \theta_\ell / \ell(x^*)$ pour des quantiles élevés de $\ell(X^*)$ pour une copie indépendante X' de X, on peut interpréter $\ell(x^*)/\theta_\ell$ comme la période de retour d'un événement extrême x^*.

Vérification et analyse de la condition de stabilité asymptotique

La condition d'appartenance au domaine d'attraction d'un max-stable dans (3.3) est essentielle pour l'utilisation des processus de Pareto. En pratique, cela nécessite de pouvoir supposer être en situation de dépendance asymptotique, au moins pour des petites distances/petits écarts, dans l'espace et dans le temps. En pratique, il convient donc de vérifier qu'il est raisonnable de faire cette hypothèse de dépendance asymptotique. Plusieurs approches sont possibles. L'une d'entre elles, consiste à évaluer l'indépendance entre les quantités observées suivantes : $\ell(X^*)$ et $X^*/\ell(X^*)$. Une autre façon de faire, plus classique, consiste à étudier empiriquement les paramètres de dépendance extrémale. Les coefficients extrémaux bivariés fournissent un résumé de la dépendance extrémale en fonction des distances dans l'espace et dans le temps en étant calculés par paires. Nous considérons d'abord la fonction coefficient extrémal dans l'espace $\theta^{spa}(h)$ pour mesurer la dépendance extrémale entre des sites séparés par la distance $||h||$ à un instant donné, et ensuite la fonction de coefficient extrémal dans le temps $\theta^{tim}(k)$ pour mesurer la dépendance extrême pour un écart temporel k à un site donné. Nous estimons $\theta^{spa}(h)$ en utilisant les observations $(X(s,t_i), X(s+h,t_i))$, et nous estimons $\theta^{tim}(k)$ à partir des paires observées $(\max_{s\in\mathcal{S}} X(s,t_i), \max_{s\in\mathcal{S}} X(s,t_i+k))$.

Certaines études empiriques menées sur des données climatiques montrent que la dépendance extrémale peut s'affaiblir quand la force de l'événement augmente c'est-à-dire en allant vers des valeurs de plus en plus extrêmes (Davison *et al.* , 2013; Thibaud *et al.* , 2013; Opitz *et al.* , 2015; Huser & Wadsworth, 2019; Le *et al.* , 2018; Tawn *et al.* , 2018). Il est alors possible que la force de la dépendance finisse par se stabiliser mais à des niveaux très élevés et non observés ou encore il est possible d'être dans la situation d'indépendance asymptotique. Avec des échantillons de taille finie, il faut reconnaitre que nous ne pouvons pas vérifier cela avec certitude mais simplement identifier ce qui peut sembler raisonnable.

3.4 Méthodologie proposée pour simuler des événements spatio-temporels extrêmes

Dans cette section, nous décrivons l'algorithme pour l'extraction d'épisodes spatio-temporels extrêmes (section 3.4.1) ainsi que la procédure générale pour simuler de nouveaux scénarios spatio-temporels (section 3.4.2). Une interprétation probabiliste d'une telle procédure est donnée section 3.4.3. Dans la suite et sans perte de généralité, nous utilisons la même notation pour l'observation du processus $X(s,t)$ et pour le processus stochastique lui-même.

3.4.1 Sélection d'épisodes extrêmes

L'algorithme 1 décrit la procédure d'extraction des épisodes extrêmes à partir des données standardisées X^*. Pour ce faire, il convient de fixer au préalable la fonctionnelle de coût ℓ. Cette dernière quantifie la force d'un événement basé en s et t faisant intervenir son voisinage $\mathcal{N}(s,t)$ qu'il convient également de définir. Nous devons également choisir un seuil u pour cette fonctionnelle suffisamment élevé pour qu'il soit raisonnable de se baser sur les résultats asymptotiques

présentés précédemment. La première étape de l'algorithme est de calculer les valeurs ℓ pour chaque voisinage $\mathcal{N}(s,t)$. Le premier épisode sélectionné correspond au voisinage $\mathcal{N}(s_1,t_1)$ pour lequel $\ell_{s,t}$ atteint la valeur maximale ℓ_1. Le second épisode sélectionné correspond à la valeur maximale de $\ell_{s,t}(X^*)$ une fois bien sûr retirés les temps correspondant aux épisodes préalablement sélectionnés (ici le premier) et autres pas de temps additionnels pour éviter les intersections et garantir l'indépendance des épisodes extraits. Cette procédure est ensuite itérée tant qu'on n'a pas atteint le nombre d'épisodes à extraire ou que la force des épisodes, quantifiée au moyen de ℓ, est supérieure au seuil u.

Algorithme 1 : Algorithme de sélection des épisodes extrêmes définis sur des voisinages spatio-temporels $\mathcal{N}(s,t)$. En étape 8, au lieu d'extraire seulement le voisinage extrême $\mathcal{N}(s_i,t_i)$, on peut souhaiter extraire l'ensemble du domaine spatial de l'étude $\mathcal{N}(t_i)\times\mathcal{S}$.

Entrées :

- $\{X^*(s,t), s\in\mathcal{S}, t\in\mathcal{T}\}$, les observations spatio-temporelles à l'échelle standardisée ;
- $\mathcal{S}'\subseteq\mathcal{S}$ les sites d'intérêt et $\mathcal{T}'\subseteq\mathcal{T}$ les pas de temps d'intérêt ;
- m' le nombre maximum d'épisodes extrêmes à sélectionner ;
- u le seuil sur $\ell_{s,t}(X^*)$ pour la sélection d'épisodes extrêmes ;
- $\delta>0$ la durée des épisodes extrêmes définis sur les voisinages temporels $\mathcal{N}(t)=[t-(\delta-1),t]$;
- $\beta\geq 0$ une marge au niveau des pas de temps pour assurer des épisodes extrêmes indépendants. On définit alors des voinages temporels étendus $\mathcal{N}_{\text{buffer}}(t)=[t-(\delta-1)-\beta, t+(\delta-1)+\beta]$;
- $\mathcal{N}(s)$ le voisinage spatial pour $s\in\mathcal{S}'$, tel que $\mathcal{N}(s,t)=\mathcal{N}(s)\times\mathcal{N}(t)$.

Sorties :

- m : le nombre d'épisodes extrêmes sélectionnés ($m\leq m'$) ;
- $\left\{X^*_{[1]}, X^*_{[2]}, \ldots, X^*_{[m]}\right\}$, $\{s_1,s_2,\ldots,s_m\}$, $\{t_1,t_2,\ldots,t_m\}$, $\{\ell_1,\ell_2,\ldots,\ell_m\}$: collection d'épisodes extrêmes ; sites et temps d'observation ; valeurs agrégées (forces) des épisodes extrêmes.

1 **début**
2 Poser $\mathcal{I}=\mathcal{T}'$.
3 Calculer $\ell_{s,t}(X^*)$ pour tout $t\in\mathcal{T}', s\in\mathcal{S}'$ avec $\mathcal{N}(s,t)\subset\mathcal{S}\times\mathcal{T}$.
4 $i\leftarrow 1$.
5 **tant que** $i\leq m'$ *et* $\max_{s\in\mathcal{S}',t\in\mathcal{I}}\ell_{s,t}(X^*)>u$ **faire**
6 $(s_i,t_i)\leftarrow\arg\max_{t\in\mathcal{I},s\in\mathcal{S}'}\ell_{s,t}(X^*)$
7 $\ell_i\leftarrow\ell_{s_i,t_i}(X^*)$
8 $X^*_{[i]}\leftarrow\{X^*(s',t'),\ (s',t')\in\mathcal{N}(s_i,t_i)\}$
9 $\mathcal{I}\leftarrow\mathcal{I}\setminus\mathcal{N}_{\text{buffer}}(t_i)$
10 $i=i+1$
11 $m\leftarrow i-1$
12 **retourner** m, $\left\{X^*_{[1]}, X^*_{[2]}, \ldots, X^*_{[m]}\right\}$, $\{s_1,s_2,\ldots,s_m\}$, $\{t_1,t_2,\ldots,t_m\}$, $\{\ell_1,\ell_2,\ldots,\ell_m\}$

3.4.2 Méthode de simulation semi-paramétrique

Pour générer de nouveaux scénarios extrêmes spatio-temporels, nous procédons de la façon suivante :

1. **Standardisation** : Estimer $\gamma(s,t)$, $\sigma(s,t)$ et $\mu(s,t)$ dans (3.6), et noter $X^* = \{T(X(s,t))\}_{s\in\mathcal{S},t\in\mathcal{T}}$ le processus ainsi standardisé (3.5).
2. **Sélection d'épisodes extrêmes** : Utiliser l'algorithme 1 pour extraire une collection m d'épisodes extrêmes $X^*_{[i]}$, $i = 1,\dots,m$.
3. **Amplification** : Échantillonner R_i, $i = 1,\dots,m$ selon une distribution de Pareto de paramètre de forme 1 et d'échelle $\alpha > 0$, i.e., $\mathbb{P}(R_i > x) = \alpha/x$, $x \in [\alpha,\infty)$, et générer des épisodes extrêmes modifiés comme suit

$$V_i(s,t) = R_i \frac{X^*_{[i]}(s,t)}{\ell_i} = R_i Y_i(s,t),\ (s,t) \in \mathcal{N}(s_i,t_i). \tag{3.9}$$

4. **Transformation à l'échelle d'origine** : Les épisodes extrêmes amplifiés sont re-transformés pour revenir à leur échelle d'origine par $W_i(s,t) = T^{\leftarrow}(V_i(s,t))$, $(s,t) \in \mathcal{N}(s_i,t_i)$.

3.4.3 Interprétation de la procédure

Selon la section 3.3.2, la procédure décrite en section 3.4.2 génère de nouvelles réalisations V_i de processus de Pareto spatio-temporel sur $\mathcal{N}(s_i,t_i)$ pour chaque épisode i, $i = 1,\dots,m$.

Comme expliqué en section 3.3.3, de $\mathbb{P}(\ell(X^*) > x) \sim \theta_\ell/x$ pour x grand, nous en déduisons qu'il est possible d'interpréter $\ell(x^*)/\theta_\ell$ comme la période de retour d'un événement extrême x^*. Par construction, les scénarios amplifiés V_i ont les mêmes motifs spatiaux de variabilité que ceux présents dans les valeurs observées mais correspondent, avec un choix approprié de α, à des périodes de retour plus longues. En effet, après amplification, et comme les réalisations de R_i sont supérieures au paramètre α, les périodes de retour des nouvelles réalisations V_i sont toujours supérieures à α/θ_ℓ. De plus, plus la valeur choisie pour α sera grande, plus longue sera la période de retour des épisodes générés et il suffira de considérer α plus grand que $\ell(x^*)$ pour s'assurer que les événements générés soient plus extrêmes que x^*.

De plus, en suivant des arguments similaires à ceux utilisés dans **[G6]**, il est possible de montrer qu'après normalisation, les épisodes générés W_i ont approximativement la même distribution que les épisodes observés X leur servant de base. Notre procédure peut alors facilement s'interpréter comme une simple élévation du seuil à un niveau de retour marginal correspondant à la période de retour de $X(s,t)$ multipliée par un coefficient r_i/ℓ_i. Ce coefficient sera supérieur à 1, correspondant bien à une amplification, si $\alpha > \ell_i$.

3.5 Simulation d'épisodes pluvieux extrêmes

La méthodologie proposée en section précédente est mise en œuvre pour produire des scénarios spatio-temporels de précipitation dans le sud de la France. Nous nous appuyons sur des données

de réanalyses horaires sur une grille de résolution 1km. Ce sont des données de réanalyses au sens où elles sont construites à partir de données radar et sont enrichies de données observées aux stations (Tabary *et al.* , 2012). La grille couvre une zone de 133,2km × 104,3 km et la période 1997-2007. Il s'agit de données Météo France.

Respectant les questions et choix à faire soulevés en section 3.3.3, nous avons proposé une transformation des marges appropriée à notre cas d'étude décrite ci-après, nous avons considéré deux fonctionnelles de coût $\ell_{s,t}^{(1)}$ et $\ell_{s,t}^{(2)}$ et analysé la dépendance extrémale comme expliqué en section 3.3.3.

Afin de préciser la distribution T, nous proposons l'utilisation d'une distribution continue G présentant une masse en 0 pour l'absence de précipitation, une densité uniforme sur $(0, x_0)$ et une Pareto standard au delà de x_0 avec $x_0 > 1$.

Pour ce qui est des fonctionnelles de coût choisies, $\ell_{s,t}^{(1)}$ est une moyenne spatio-temporelle, c'est-à-dire la valeur moyenne de $X^*(s,t)$ sur des voisinages spatio-temporels $\mathcal{N}(s,t) = \mathcal{N}(s) \times \mathcal{N}(t)$. Dans l'espace nous avons considéré un disque de rayon 15km centré en s ($\mathcal{N}(s) = \{s' \in \mathcal{S} \mid \|s - s'\| \leq 15 \text{ km}\}$) et pour le temps, les $11h$ précédant t et t correspondant à une durée de $\delta = 12$ heures ($\mathcal{N}(t) = \{t' \in \mathcal{T} \mid |t - t'| \leq 12 \text{ heures}\}$). La seconde fonctionnelle de coût utilisée est le maximum spatio-temporel. En d'autres termes, il s'agit du max spatial sur toute la zone ($\mathcal{N}(s) = S$ et $\ell_{s,t}^{S} = max$). Le voisinage temporel reste ici le même que précédemment et $\ell_T = max$. Dans l'analyse de la dépendance extrémale menée, les fonctions coefficients extrémaux dans l'espace et dans le temps ont été estimées et indiquent qu'il est tout à fait raisonnable de supposer la dépendance asymptotique quelle que soit la distance entre deux sites de la zone. Ils indiquent que cette hypothèse est également raisonnable dans le temps pour des épisodes d'une durée d'une dizaine d'heures maximum. Ce dernier résultat est en accord avec le choix de $\delta = 12$ heures mentionné précédemment et nous empêche de considérer des épisodes plus longs.

La méthodologie proposée a été mise en application en suivant la procédure d'extraction et d'amplification de la section 3.4. Le paramètre β a été fixé égal à 1 pour séparer les épisodes extraits d'au moins 1 heure et le seuil u est choisi comme un quantile élevé (95% ou 98%) de $\ell_{s,t}^{(j)}$, $j = 1, 2$ après une étape de pré-traitement visant à réduire la proportion de 0. Le paramètre d'échelle α de la Pareto de la variable R_i a été choisi comme le double de la valeur des forces observées $\ell_i^{(1)}(X^*)$ et $\ell_i^{(2)}(X^*)$. Nous avons extrait $m = m' = 6$ épisodes selon chacune des 2 fonctionnelles. Sur les 6 épisodes extraits, 4 épisodes pluvieux ont été sélectionnés par les deux fonctionnelles. Les autres épisodes sont plus spécifiques et peuvent correspondre soit à de très fortes pluies mais de façon très localisée dans le temps ou dans l'espace ou à des pluies modérées mais très soutenues dans l'espace et dans le temps. Je fais le choix dans ce document de présenter en Figure 3.1 les résultats du processus d'amplification du 4ème et du 6ème épisode extrait pour la moyenne spatio-temporelle $\ell_{s,t}^{(1)}$ et renvoie le lecteur à **[G20]** pour d'autres illustrations. Le 4ème épisode correspond à un épisode qui n'a pas été sélectionné par la seconde fonctionnelle alors que le 6ème correspond au 5ème épisode sélectionné selon la deuxième fonctionnelle $\ell_{s,t}^{(2)}$. Le paramètre α de la Pareto est pris égal au double de la valeur maximum de ℓ_i, $i = 4, 6$, de sorte qu'on obtient des épisodes extrêmes amplifiés de période de retour au moins deux fois plus longue que celle de l'épisode d'origine.

Nous renvoyons le lecteur à l'article **[G20]** pour une analyse de risque qui vise à explorer les différences dans les épisodes extrêmes amplifiés qui peuvent être imputées au choix des fonc-

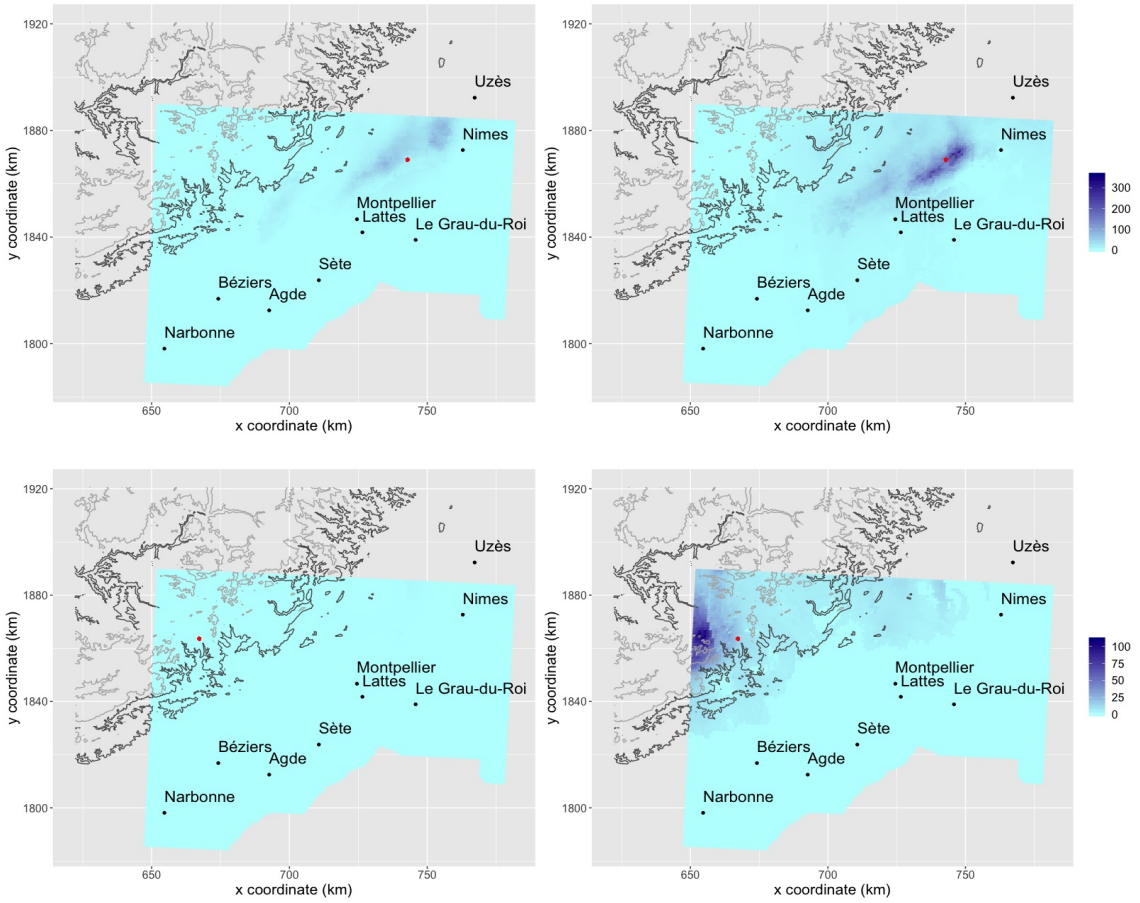


FIGURE 3.1 – Données d'origine de précipitation $X(s,t)$ (à gauche) et épisodes amplifiés $W(s,t)$ (à droite) basés sur la moyenne spatio-temporelle $\ell_{s,t}^{(1)}$. En haut t =2002-09-08, en bas t =2001-07-06. Les points rouges indiquent le site s_i pour lequel on a observé la valeur maximum ℓ_i durant l'épisode sélectionné, $i = 4, 6$. Les contours en gris et noir indiquent l'altitude.

tionnelles de coût et du seuil inférieur fixe (c'est-à-dire le paramètre d'échelle α de la variable Pareto R_i). Plus précisément, les 3 épisodes les plus extrêmes, selon les deux fonctionnelles, sont amplifiés en utilisant pour R_i les quartiles de la loi de Pareto de paramètre d'échelle α et de forme 1 avec α variant du simple au quadruple des valeurs maximales de ℓ sur l'épisode d'origine. Les comparaisons se font sur la base du calcul de quantile extrême et de l'espérance conditionnelle de queue (CTE en anglais pour Conditional Tail Expectation).

3.6 Perspectives

Cette contribution de simulation d'épisodes extrêmes spatio-temporels s'inscrit dans le formalisme des processus de Pareto que nous avons alors réécrit dans un cadre spatio-temporel. L'approche de simulation développée est une approche semi-paramétrique. Elle est totalement non-paramétrique pour ce qui est de la structure de dépendance et paramétrique pour les marginales s'appuyant alors sur la théorie univariée des extrêmes. Dans un cadre spatial, Thibaud

& Opitz (2015) et De Fondeville & Davison (2018) adoptent une approche paramétrique, les derniers analysant les pluies extrêmes en Floride en ajustant un processus de Pareto basé sur les processus log-Gaussiens. Une composante importante de l'approche que nous avons développée est la réflexion autour de la fonctionnelle de coût définie sur une fenêtre spatio-temporelle glissante. Cela permet de caractériser les épisodes extrêmes comme des épisodes dont le "coût" excède un certain seuil. On est alors capable de construire de nouveaux épisodes ayant une période de retour plus longue que ceux observés. Notre approche, comme celle de De Fondeville & Davison (2018), définit les excès sur les données transformées. Il est parfois plus naturel et facilement interprétable de chercher à les définir sur les données d'origine. Moyennant un indice de queue constant sur l'espace, c'est ce que permet notamment l'approche très récente de De Fondeville & Davison (2020) qui proposent une extension fonctionnelle de la distribution de Pareto généralisée généralisant les travaux de Dombry & Ribatet (2015).

Notre approche a été récemment mise en œuvre sur des données de précipitation, l'objectif étant la création d'épisodes pluvieux intenses en zone méditerranéenne. Il s'agit en effet d'une région exposée à certains risques naturels causés par des événements météorologiques extrêmes tels que des périodes de fortes précipitations (en termes de durée et/ou d'intensité).

Une des perspectives à l'utilisation d'une méthodologie de simulation d'épisodes extrêmes est l'étude du risque inondation en milieu urbain. Pour étudier le risque d'inondation dans les zones urbaines, des modèles d'écoulement peuvent être utilisés. Ces modèles d'écoulement doivent être conditionnés par des forçages, c'est-à-dire des variables telles que les précipitations que l'on met en entrée des modèles déterministes. L'équipe Inria LEMON développe de nouveaux modèles d'écoulement et, de par ma contribution qui fait l'objet de ce chapitre, propose donc également de simuler stochastiquement des forçages extrêmes pluviométriques en contrôlant leur force. L'évaluation des impacts de ces épisodes extrêmes sur le risque d'inondation consisterait donc à se donner un catalogue d'épisodes pluvieux extrêmes dont on fixerait un certain nombre de caractéristiques et d'alimenter avec ces derniers un modèle d'écoulement en milieu urbain. Nous récupérons alors en sortie des informations comme les hauteurs et vitesse d'eau dans la ville à partir desquelles des mesures de risque peuvent se calculer. Mener une étude d'impacts revient alors à étudier le lien entre les caractéristiques des forçages et celles des mesures de risque déduites en sortie. Il est alors clair que déterminer les mesures de risque à considérer fait également partie de la tâche à mener et qu'il conviendra de considérer des mesures de risque multivariées, les indicateurs de risque d'inondation étant généralement dérivés en combinant différentes variables hydrauliques. Cette perspective fait l'objet de l'axe 2 du projet de recherche (voir section 4.2 du chapitre 4).

Une autre perspective à ce travail de simulation d'épisodes pluvieux intenses est l'intégration de ces derniers dans des simulations de longues séries d'événements pluvieux courants et de temps sec. Loin d'être immédiate, cette combinaison de scénarios extrêmes et non extrêmes nécessite une modélisation réaliste des transitions entre périodes normales et extrêmes. Naveau *et al.* (2016) ont proposé un modèle statistique pouvant servir de générateur de pluie dans un cadre univarié. Dans un travail en cours **[P-2]**, nous tirons profit de ces travaux en univarié combinés à une approche de modélisation hiérarchique comme dans **[G3]** pour modéliser la dynamique temporelle pour toute la gamme de valeurs (zéros et extrêmes inclus) et ce sans avoir à définir un seuil a priori. Cette perspective est également reprise dans le projet de recherche (voir section 4.1 du chapitre 4).

Chapitre 4

Projet de recherche

Mon projet de recherche s'intitule **Modélisation stochastique et analyse statistique de processus climatiques et littoraux extrêmes**. J'ai fait le choix de le structurer en **3 axes**.

Le premier axe, *Modélisation statistique d'événements extrêmes*, présenté en section 4.1, se trouve dans la parfaite continuité de mes travaux récents en modélisation spatiale et spatio-temporelle pour les extrêmes. J'y présente plus en détail trois travaux en cours dont deux ont également été mis en avant dans les perspectives des chapitres 2 et 3. Le second axe, *Étude du risque inondation en milieu urbain* (section 4.2), est articulé en six étapes dont deux d'entre elles s'appuient sur les travaux théoriques développés dans le premier axe. Enfin la section 4.3, *Modélisation statistique de phénomènes complexes*, traite de l'apport de la théorie des valeurs extrêmes pour la prédiction des distributions d'espèces en écologie.

4.1 Modélisation statistique d'événements extrêmes

Ce premier axe se positionne dans le prolongement de mes derniers travaux de recherche. Il s'agit de développer, proposer, étudier et mettre en œuvre des modèles adaptés à la présence de valeurs extrêmes. Je me place dans **un cadre multivarié (A), temporel (B) et/ou spatial (C)**, l'objectif étant de tendre vers la prise en compte de deux voire trois aspects simultanément.

Se placer dans les contextes sus-mentionnés représente une première difficulté et nécessite de tenir compte des dépendances complexes associées. J'énumère ci-dessous trois autres verrous notés **(V1)**, **(V2)** et **(V3)** auxquels je porte une attention particulière. Dans cet axe, je cherche à combiner la prise en compte de certains d'entre eux.

> **(V1)** Présence **d'indépendance asymptotique.** L'objectif est de prendre en compte l'indépendance asymptotique de la façon la plus flexible possible. Par exemple, on souhaite le faire éventuellement de manière partielle c'est-à-dire sur certaines composantes uniquement (spatiale et non temporelle par exemple, entre certaines composantes d'un vecteur et pas d'autres...). Cela me place dans la continuité des développements présentés dans le chapitre 2.

(V2) Prise en compte **d'une non-stationnarité spatiale et/ou temporelle de la structure de dépendance**. Les zones d'influence en terme de dépendance spatiale pourraient par exemple être différentes selon les régions, ce qui peut être très utile dans le sud de la France où il faut composer avec notamment les Cévennes et une zone littorale.

(V3) Prise en compte simultanée d'événements extrêmes et d'événements communs. L'objectif est d'être en capacité **de combiner dans l'espace et dans le temps des simulations d'événements extrêmes et des simulations d'événements courants**. À titre d'exemple, si l'on s'intéresse aux pluies, on aimerait intégrer des événements extrêmes dans de longues chroniques incluant des pas de temps sec et des pluies moyennes.

Pour illustrer les travaux qui pourraient être menés dans cet axe, je présente ci-dessous trois travaux **[P-1]**, **[P-2]** et **[P-3]** qui sont actuellement en préparation.

Dans **[P-1]**, repartant du fait que la distribution de Pareto est un mélange d'une loi exponentielle avec une loi Gamma, nous étudions dans une collaboration avec J.N. Bacro, C. Gaetan et T. Opitz le ratio par composante de deux vecteurs aléatoires de distributions marginales exponentielle et Gamma. Nous caractérisons les propriétés de dépendance extrémale des distributions multivariées obtenues, les marginales étant par construction Pareto. Nous définissons alors un cadre de modélisation flexible pour les extrêmes **multivariés** (**cadre A** ci dessus), permettant la dépendance asymptotique entre certaines composantes et l'indépendance asymptotique entre certaines autres (répondant alors à **(V1)**). De premiers résultats prometteurs ont été obtenus sur un jeu de données de pollution à Milan.

Dans un cadre univarié, Naveau *et al.* (2016) ont proposé un modèle statistique pouvant servir de générateur de pluie. Dans une nouvelle collaboration engagée lors de la venue de P. Naveau en avril 2018 à Montpellier et également avec T. Opitz, nous tirons profit de ces travaux combinés à une approche de modélisation hiérarchique comme proposée dans **[G3]** afin de proposer une approche de simulation pour toute la gamme de valeurs. Une force de ce travail est de ne pas avoir à définir un seuil a priori (voir **[P-2]**). L'approche est alors **temporelle** (**cadre B** ci-dessus) et s'attaque au verrou **(V3)**. Elle pourrait également permettre une non-stationnarité temporelle de la dépendance **(V2)**.

Dans **[G7]**, un modèle spatial hybride a été proposé permettant différents types de dépendance extrémale en fonction des distances considérées. Une extension de ce modèle faisant l'objet d'un travail en préparation avec J.N. Bacro, J. Carreau et C. Gaetan **[P-3]** permettra de considérer ces différentes dépendances également selon la zone spatiale assurant alors une non-stationnarité spatiale de la structure de dépendance. Le cadre sera alors **spatial (cadre C)** et nous nous intéresserons aux verrous **(V1)** et **(V2)**. L'idée est de faire dépendre le paramètre de mélange de la position spatiale comme cela a été fait dans **[G1]**. Dans **[G1]**, disponible en annexe D et réalisé en collaboration avec J. Carreau, l'accent a été mis sur la prise en compte de la non-stationnarité non pas dans les intensités mais dans la structure de dépendance spatiale. Plus précisément, nous avons généralisé dans un cadre spatial le mélange de deux copules de Gumbel. Cette approche permet d'obtenir les distributions en tout site de la zone étudiée, qu'ils correspondent à des stations ou à des points quelconques de l'espace. Par ailleurs, l'inférence des paramètres de la dépendance a été réalisée de façon originale, nous appuyant notamment sur une technique ABC (Approximate Bayesian Computation) peu utilisée pour les modèles d'extrêmes

spatiaux jusqu'à présent. Cela a notamment permis de conduire une application sur des maxima annuels de pluie dans la région des Cévennes. S'appuyant aussi sur **[G1]**, une version spatio-temporelle de **[G7]** devrait également permettre de construire de la non-stationnarité temporelle.

Les travaux amorcés ou envisagés dans cet axe s'inscrivent parfaitement dans mes projets actuels et notamment dans le projet LEFE intitulé FRAISE, que je porte depuis janvier 2019, fédérant un certain nombre de chercheurs de différentes disciplines. Un des objectifs spécifiques visés dans FRAISE et qui constitue encore un véritable défi (Ailliot *et al.* , 2015) concerne précisément la génération de scénarios spatio-temporels de forçages de précipitations qui intégrerait des extrêmes. Une partie des travaux de cet axe s'inscrit également dans le cadre du projet PHC Utique (2019-2021) avec la Tunisie intitulé AMANDE (Approches stochastiques et seMipAramétriques combinées à la télédétectioN pour l'étude du stress hyDriquE) porté par Julie Carreau (IRD) et dont je suis partenaire.

4.2 Étude du risque inondation en milieu urbain

Un des risques naturels les plus destructeurs, créant des dommages matériels et humains considérables, est le phénomène des crues éclair. Ces crues peuvent être déclenchées par des pluies intenses localisées de quelques heures ou par des pluies de plus longues durées avec des intensités modérées. Pour étudier le risque inondation en milieu urbain, on peut avoir recours à des modèles à base physique qui permettent de caractériser et de simuler les écoulements à partir des connaissances sur les processus. Ces modèles déterministes doivent être conditionnés par des forçages, i.e, des données en entrée telles que les épisodes pluvieux. Construire ces scénarios de forçages au plus près du réel représente donc un enjeu primordial. Les approches stochastiques permettent de simuler des scénarios de forçages de façon aléatoire. Pour construire ces scénarios de forçages, deux étapes peuvent être identifiées. La première **(E1)** est la **modélisation de processus spatio-temporels extrêmes de pluie et le développement de méthodes de simulation associées** et la seconde **(E2)** concerne la **combinaison, dans la simulation, d'événements extrêmes et non extrêmes** dans le but de construire des chroniques de pluies aussi réalistes que possible intégrant des pluies extrêmes, des pluies courantes et des périodes de temps sec.

Ces premières étapes **(E1)** et **(E2)** s'appuient complètement sur les travaux de l'axe 1 qui seront réalisés au maximum dans cette direction. La compréhension de la variabilité spatiale et temporelle des pluies pouvant générer des crues éclair représente un enjeu majeur. Ces connaissances sont essentielles pour construire les méthodes stochastiques de simulation de scénarios intégrant des champs extrêmes spatio-temporels réalistes de pluies. La modélisation de la structure spatio-temporelle des pluies extrêmes comme les épisodes cévenols devra se faire en gardant à l'esprit l'importance de l'interprétation physique de données simulées selon de tels modèles. C'est pourquoi l'accent sera mis sur les approches basées sur les dépassements permettant une interprétation des champs simulés comme des événements.

Se donner des mesures de risque appropriées pour ensuite pouvoir les exploiter afin d'évaluer les impacts potentiels d'épisodes (intégrant des) extrêmes de pluie simulés stochastiquement mis en entrée/forçage du modèle d'écoulement fait l'objet de la troisième étape **(E3)** de cet axe.

L'estimation du risque d'inondation est complexe, du fait à la fois des multiples aspects que celui-ci peut revêtir et de la non-linéarité des processus. Pour les personnes, le risque obéit en général à des comportements à seuils. Quand certaines variables hydrodynamiques (hauteur d'eau, vitesse) ou leur combinaison dépassent certaines valeurs limites, les piétons sont susceptibles d'être renversés et/ou emportés par le courant. Il n'existe que peu de littérature sur les mesures de risque multivariées ou spatiales pour l'étude du risque inondation. Pourtant, quantifier un risque pouvant dépendre de plusieurs quantités d'intérêt évoluant elles-mêmes dans l'espace et dans le temps est capital. Un travail est initié avec T. Opitz et F. Palacios-Rodriguez **[P-4]** sur la proposition et l'étude d'une mesure de risque multivariée faisant intervenir des moyennes ou des quantiles (ou autre fonction) des composantes d'un vecteur conditionnellement à ce qu'une fonctionnelle du vecteur dépasse un certain seuil. Un autre travail est en cours sur l'exploitation des modèles issus de la thèse de N. Dalhoumi pour l'estimation d'ensembles à risque, l'idée étant dans le cadre le plus simple d'identifier par exemple toutes les combinaisons de valeurs hauteurs/vitesses d'eau qui conduiraient à un risque supérieur à un niveau fixé. Il s'agit d'un travail **[P-5]** en collaboration avec F. Palacios Rodriguez, J.N. Bacro et E. Di Bernardino.

Une fois des mesures appropriées identifiées et/ou définies, nous serons en capacité dans une quatrième étape **(E4)** de mener des études d'impacts hydrologiques en évaluant justement les impacts de forçages de pluies c'est-à-dire d'épisodes extrêmes que nous aurions générés. Il s'agit d'une perspective détaillée au chapitre 3 en section 3.6.

D'autres volets statistiques sont utiles pour mener cet axe de manière complète. En particulier des techniques de downscaling statistique pourraient intervenir à 2 niveaux. D'une part, dans un contexte d'inondation urbaine, les modèles d'écoulement nécessitent en forçage des simulations de précipitation ayant une résolution spatiale inférieure au pixel radar ($1km^2$). Depuis 2019, un observatoire urbain est mis en place à Montpellier par le laboratoire HydroSciences Montpellier et nous y contribuons par l'intermédiaire du projet FRAISE. Cet observatoire crée un réseau dense d'une vingtaine de pluviographes sur le campus Triolet de l'Université de Montpellier et alentours. Les données acquises nous permettront de proposer des forçages à la bonne résolution spatiale en mettant en œuvre ou en développant des techniques de descente d'échelle **(E5)**. D'autre part, les modèles d'écoulement 2D en zone urbaine requièrent des temps de calculs trop importants pour fournir les simulations nécessaires au calcul d'indicateurs servant aux systèmes de premières alertes. Une solution consiste à considérer un modèle ayant un maillage à plus Basse Résolution (BR) spatiale qui approche le comportement du modèle 2D Haute Résolution (HR). Ainsi, les modèles à porosité développés dans l'équipe LEMON pour les inondations urbaines (Guinot, 2012; Guinot *et al.* , 2018) sont de 100 à 1000 fois plus rapides que les modèles « shallow water » habituellement utilisés. Néanmoins, si l'on veut s'appuyer sur les modèles à porosité dans l'étude de risque il faut être capable de reconstituer les champs hydrodynamiques à échelle fine à partir des résultats des modèles à porosité. Cela constitue l'étape **(E6)**.

Très clairement cet axe s'appuie de nouveau sur le projet LEFE FRAISE. Il est également totalement imbriqué dans les objectifs du projet LEMON. Pour les aspects mesures de risque, je participe à deux projets qui sont en cours d'évaluation. Le premier est un projet JCJC porté par T. Laloé, intitulé "MaChine Learning And Risk EvaluatioN" (MCLAREN) dans lequel j'interviens en tant qu'experte sur la thématique des valeurs extrêmes. Le second a été soumis par F. Palacios au programme espagnol Becas Leonardo 2020 de la fondation BBVA. Il s'agit principalement de financements de missions entre l'Espagne et la France pour faciliter nos échanges.

4.3 Modélisation statistique de phénomènes complexes

Je souhaite dans ce troisième axe m'intéresser à la modélisation et à la prédiction des distributions des espèces, ce qui représente un enjeu majeur pour préserver les écosystèmes naturels en particulier face au changement climatique et à l'augmentation des pressions humaines. Classiquement, l'abondance des espèces est supposée distribuée selon une loi de Poisson, dont l'intensité dépend de caractéristiques environnementales. Toutefois, en raison de différents facteurs (dispersion limitée, compétition entre espèces, etc), les données d'abondances présentent une surdispersion qui se caractérise soit par un excès de zéros, soit par des valeurs extrêmes soit les deux simultanément. Cette surdispersion viole la propriété d'égalité entre l'espérance et la variance d'une loi de Poisson. Les modèles mélanges de lois de Poisson forment une solution élégante pour gérer la question de la surdispersion (Karlis & Xekalaki, 2005). Cette construction s'interprète aussi d'un point de vue bayésien, la loi de mélange étant alors la loi *a priori* de l'intensité de la loi de Poisson. Les objectifs sont d'étudier et de caractériser les lois de Poisson en mélange en fonction des propriétés de la loi a priori de l'intensité, en se focalisant non seulement sur leur comportement en zéro, mais aussi sur la structure des queues de distribution (apport de la théorie des valeurs extrêmes et étude du cas discret). D'un point de vue pratique, cela permettra de guider le choix de la loi de mélange selon les observations. Sur cette thématique, et dans le cadre de l'ANR GAMBAS, je co-encadre avec F. Mortier (CIRAD) et J. Pehardi (UM), S. Valiquette, en stage de M2 biostatistique (février-août 2020). Je démarre également le co-encadrement d'une thèse à partir de septembre 2020 sur le développement mathématique de modèles de distributions d'espèces. La première étape consistera à consolider les résultats obtenus au cours du stage puis à les généraliser au cas multivarié pour modéliser plusieurs espèces simultanément.

4.4 Conclusion

Comme sus-mentionné, mes 3 axes de recherche s'inscrivent dans une dynamique de projets dont certains sont extrêmement structurants. En particulier les axes 1 et 2 rejoignent des objectifs identifiés dans les deux projets LEFE que je porte, CERISE et FRAISE (2016-2021). Ils sont également en parfaite adéquation avec le projet de l'équipe Inria LEMON pour lequel j'ai activement participé à la rédaction de la feuille de route associée. L'équipe LEMON a été officiellement créée le 1er janvier 2019 et compte aujourd'hui 5 permanents (1 chargé de recherche Inria et 4 enseignants-chercheurs répartis sur 2 UMR que sont HydroSciences Montpellier et l'IMAG). En quelques mots, l'objectif général de LEMON est de développer des méthodes mathématiques et computationnelles pour la modélisation de processus littoraux et environnementaux. Les outils mathématiques utilisés sont à la fois déterministes mais aussi probabilistes et statistiques. Alors que les 4 autres permanents de l'équipe sont principalement impliqués sur les aspects déterministes, mon travail s'inscrit dans la composante aléatoire du projet de l'équipe et couvre les aspects tant probabilistes que statistiques.

La majorité des perspectives de recherche que j'ai introduites sont présentées avec des applications liées au risque inondation et il s'agit clairement de l'un de mes champs d'applications privilégiés dans les années à venir. Néanmoins, je poursuivrai mon investissement sur la modélisation des vagues. En particulier je fais partie du Défi Inria SURF (Sea Uncertainty Repre-

sentation and Forecast) de structuration des contributions en modélisation océanique. Les Défis Inria fonctionnent comme des équipes-projets inter-équipes et sont mis en place pour permettre une organisation autour de sujets de recherche jugés majeurs par l'institut. Je pense aussi à ma collaboration démarrée il y a maintenant 8 ans avec Géosciences Montpellier. J'ai rejoint le réseau GLADYS, groupe de recherche sur le littoral méditerranéen qui s'institutionnalise prochainement avec la création de l'institut des plages de l'Université de Montpellier piloté par F. Bouchette. Avec des collègues de GLADYS, j'ai participé en 2018 à une mission terrain dans le cadre du projet MAUPITI HOE (projet 2017-2027) également porté par F. Bouchette. Il s'agissait de déployer du matériel autour de l'île de Maupiti permettant ainsi la collecte de données à des fins de modélisation mathématique et physique du comportement hydro-morphodynamique des atolls polynésiens. Nous avons récemment soumis un premier article **[G23]** descriptif de la campagne d'acquisition des données et des premiers résultats et constats sur la variation spatiale de la bathymétrie du récif. Notre travail se poursuit et notre objectif est d'améliorer notre compréhension et de décrire complètement les interactions entre l'hydrodynamique littorale et les bathymétries complexes. Ces informations auront vocation à être utilisées dans les modèles numériques de circulation et de propagation de la houle.

Comme l'illustrent mes travaux passés et mes projets, ma recherche est essentiellement motivée par des questions liées aux sciences de l'environnement, mais j'ai à cœur de rester ouverte à d'autres recherches, tant théoriques qu'appliquées, pouvant émaner de thématiques autres. C'est ainsi par exemple, que je m'intéresse aux extrêmes de distributions discrètes (dans l'axe 3 présenté en section 4.3) ou que j'ai pu travailler en biostatistique médicale sur des aspects méthodologiques pour le choix de dose pour la phase III des essais cliniques (co-encadrement de la thèse de J. Aouni). M'impliquer dans de nouvelles thématiques et découvrir de nouveaux aspects méthodologiques, tout comme rester proche du milieu industriel, sont importants pour moi et témoignent de la curiosité scientifique qui me caractérise. Enfin, les rencontres et collaborations sont fondamentales à mes yeux tant sur des aspects scientifiques, thématiques et humains et participent pleinement à mon épanouissement scientifique et personnel. De façon naturelle, je souhaite pour l'avenir contribuer aux développements de collaborations existantes et à venir.

Bibliographie

Abu-Awwad, A.F., Maume-Deschamps, V., & Ribereau, P. 2019. Censored pairwise likelihood-based tests for mixing coefficient of spatial max-mixture models. *Revista de Investigacion Operacional*.

Ahmed, M., Maume-Deschamps, V., Ribereau, P., & Vial, C. 2017. *A semi-parametric estimation for max-mixture spatial processes*. Preprint.

Ahmed, Manaf, Maume-Deschamps, V., Ribereau, P., & Vial, C. 2019. Spatial risk measures for max-stable and max-mixture processes. *Stochastics*, 1–16.

Ailliot, P., Allard, D., Monbet, V., & Naveau, P. 2015. Stochastic weather generators : an overview of weather type models. *Journal de la Société Française de Statistique*, **156**(1), 101–113.

Bacro, J.N., & Gaetan, C. 2014. Estimation of spatial max-stable models using threshold exceedances. *Statistics and Computing*, **24**(4), 651–662.

Barndorff-Nielsen, O.E., Lunde, A., Shepard, N., & Veraat, A.E.D. 2014. Integer-valued trawl processes : A class of stationary infinitively divisible processes. *Scandinavian Journal of Statistics*, **41**(3), 693–724.

Bortot, P., & Gaetan, C. 2014. A latent process model for temporal extremes. *Scandinavian Journal of Statistics*, **41**(3), 606–621.

Caires, S., de Haan, L., & Smith, R.L. 2011. *On the determination of the temporal and spatial evolution of extreme events*. Tech. rept. Deltares. Report 1202120-001-HYE-004 (for Rijkswaterstaat, Centre for Water Management).

Casson, E., & Coles, S.G. 1999. Spatial regression models for extremes. *Extremes*, **1**, 449–468.

Chailan, R. 2015 (Nov.). *Application of Scientific Computing and Statistical Analysis to address Coastal Hazards*. Theses, Université Montpellier.

Coles, S., Heffernan, J., & Tawn, J. 1999. Dependence measures for extreme value analyses. *Extremes*, **2**, 339–365.

Cooley, D., Nychka, D., & Naveau, P. 2007. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, **102**(479), 824–840.

Cressie, N. 1991. *Statistics for spatial data*. New York : J. Wiley & Sons.

Cressie, N., & Wikle, C.K. 2011. *Statistics for Spatio-Temporal Data*. CourseSmart Series. Wiley.

Davis, R.A., Klüppelberg, C., & Steinkohl, C. 2013a. Max-stable processes for modeling extremes observed in space and time. *Journal of the Korean Statistical Society*, **42**(3), 399–414.

Davis, R.A., Klüppelberg, C., & Steinkohl, C. 2013b. Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society*, **75**(5), 791–819.

Davison, A.C., & Gholamrezaee, M. M. 2012. Geostatistics of extremes. *Proceedings of the Royal Society London, Series A*, **468**, 581–608.

Davison, A.C., & Smith, R. L. 1990. Models for exceedances over high thresholds. *Journal of Royal Statistical Society : Series B*, **3**, 393–442.

Davison, A.C., Padoan, S. A., & Ribatet, M. 2012. Statistical modelling of spatial extremes. *Statistical Science*, **27**(2), 161–186.

Davison, A.C., Huser, R., & Thibaud, E. 2013. Geostatistics of dependent and asymptotically independent extremes. *Journal of Mathematical Geosciences*, **45**, 511–529.

De Fondeville, R., & Davison, A.C. 2018. High-dimensional peaks-over-threshold inference. *Biometrika*, **105**(3), 575–592.

De Fondeville, R., & Davison, A.C. 2020. *Functional Peaks-over-threshold Analysis.* Preprint.

de Haan, L. 1984. A spectral representation for max-stable processes. *The Annals of Probability*, **12**(4), 1194–1204.

de Haan, L., & Ferreira, A. 2006. *Extreme Value Theory : an Introduction.* New-York : Springer.

de Oliveira, T. 1962. Structure theory of bivariate extremes : extensions. *Estudos de Mathemàtica, Estatistica e Econometrica*, **7**, 165–195.

Dombry, C., & Ribatet, M. 2015. Functional regular variations, Pareto processes and peaks over threshold. *Statistics and Its Interface*, **8**(1), 9–17.

Dombry, C., Eyi-Minko, F., & Ribatet, M. 2013. Conditional simulation of max-stable processes. *Biometrika*, **100**(1), 111–124.

Dombry, C., Engelke, S., & Oesting, M. 2016. Exact simulation of max-stable processes. *Biometrika*, **103**(2), 303–317.

Engelke, S., de Fondeville, R., & Oesting, M. 2019. Extremal behaviour of aggregated data with an application to downscaling. *Biometrika*, **106**(1), 127–144.

Ferguson, T.S. 1973. A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, **1**(2), 209–230.

Ferreira, A., & de Haan, L. 2014. The generalized Pareto process ; with a view towards application and simulation. *Bernoulli*, **20**(4), 1717–1737.

Ferreira, A., de Haan, L., & Zhou, C. 2012. Exceedance probability of the integral of a stochastic process. *Journal of Multivariate Analysis*, **105**(1), 241–257.

Gaetan, C., & Grigoletto, M. 2007. A hierarchical model for the analysis of spatial rainfall extremes. *Journal of Agricultural Biological and Environmental Statistics*, **12**, 434–449.

Groeneweg, J., Caires, S., & Roscoe, K. 2012. Temporal and spatial evolution of extreme events. *Coastal Engineering Proceedings*, **1**(33), management–9.

Guinot, V. 2012. Multiple porosity shallow water models for macroscopic modelling of urban floods. *Advances in Water Resources*, **37**, 40–72.

Guinot, V., Sanders, B.F., & Schubert, J.E. 2018. Dual integral porosity shallow water model for urban flood modelling. *Advances in Water Resources*, **103**, 16–31.

Heffernan, J. E., & Tawn, J. A. 2004. A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, **66**(3), 497–546.

Huser, R., & Davison, A.C. 2014. Space-time modelling of extreme events. *Journal of the Royal Statistical Society : Series B*, **76**(2), 439–461.

Huser, R., & Wadsworth, J.L. 2019. Modeling Spatial Processes with Unknown Extremal Dependence Class. *Journal of the American Statistical Association*, **114**(525), 434–444.

Huser, R., Opitz, T., & Thibaud, E. 2017. Bridging asymptotic independence and dependence in spatial extremes using gaussian scale mixtures. *Spatial Statistics*, **21**, 166–186.

Jeon, S., & Smith, R.L. 2012. *Dependence structure of spatial extremes using threshold approach.* Tech. rept. arXiv :1209.6344.

Kabluchko, Z., Schlather, M., & de Haan, L. 2009. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability*, **37**, 2042–2065.

Karlis, D., & Xekalaki, E. 2005. Mixed Poisson Distributions. *International Statistical Review*, **73**(1), 35–58.

Le, P.D., Davison, A.C., Engelke, S., Leonard, M., & Westra, S. 2018. Dependence properties of spatial rainfall extremes and areal reduction factors. *Journal of Hydrology*, **565**, 711–719.

Ledford, A.W., & Tawn, J.A. 1996. Statistics for near independence in multivariate extreme values. *Biometrika*, **83**(1), 169 –187.

Ledford, A.W., & Tawn, J.A. 1997. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society, Series B*, **59**(2), 475–499.

Lin, T., & de Haan, L. 2001. On convergence toward an extreme value distribution in C[0,1]. *Annals of Probability*, **29**(1), 467–483.

Lindsay, B. 1988. Composite likelihood methods. *Contemporary Mathematics*, **80**, 221–239.

Naveau, P., Huser, R., Ribereau, P., & Hannart, A. 2016. Modelling jointly low, moderate and heavy rainfall intensities without a threshold selection. *Water Resources Research*, **52**, 2753–2769.

Noven, R.C., Veraart, A.E.D., & Gandy, A. 2018. A latent trawl process model for extreme values. *Journal of Energy Markets*, **11**(3), 1–24.

Oesting, M., & Stein, A. 2018. Spatial modeling of drought events using max-stable processes. *Stochastic Environmental Research and Risk Assessment*, **32**(1), 63–81.

Oesting, M., Bel, L., & Lantuéjoul, C. 2018. Sampling from a max-stable process conditional on a homogeneous functional with an application for downscaling climate data. *Scandinavian Journal of Statistics*, **45**(2), 382–404.

Opitz, T. 2013. Extremal t processes : elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, **122**, 409–413.

Opitz, T., Bacro, J.N., & Ribereau, P. 2015. The spectrogram : A threshold-based inferential tool for extremes of stochastic processes. *Electronic journal of statistics*, **9**(1), 842–868.

Padoan, S.A., Ribatet, M., & Sisson, S.A. 2010. Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, **105**(489), 263–277.

Pickands, J. 1975. Statistical inference using extreme order statistics. *The Annals of Statistics*, **3**(1), 119–131.

Reiss, R.D., & Thomas, M. 2007. *Statistical Analysis of Extreme Values*. Third edn. Basel : Birkhäuser.

Resnick, S.I. 1987. *Extreme values, Regular variation and Point Processes*. Springer Verlag.

Rootzén, H., & Tajvidi, N. 2006. Multivariate generalized Pareto distributions. *Bernoulli*, **12**(5), 917–930.

Sang, H., & Gelfand, A. 2009. Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, **16**, 407–426.

Schlather, M. 2002. Models for stationary max-stable random fields. *Extremes*, **5**, 33–44.

Schlather, M., & Tawn, J.A. 2003. A dependence measure for multivariate and spatial extreme Values : properties and inference. *Biometrika*, **90**(1), 139–156.

Serinaldi, F., Bárdossy, A., & Kilsby, C.G. 2014. Upper tail dependence in rainfall extremes : would we know it if we saw it ? *Stochastic Environmental Research and Risk Assessment*, **29**(4), 1211–1233.

Sibuya, M. 1960. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics*, **11**, 195–210.

Smith, R. L. 1990. *Max-stable processes and spatial extremes*. Unpublished paper, University of Surrey.

Tabary, P., Dupuy, P., L'Henaff, G., Gueguen, C., Moulin, L., Laurantin, O., Merlier, C., & Soubeyroux, J.-M. 2012. A 10-year (1997–2006) reanalysis of quantitative precipitation estimation over France : methodology and first results. *IAHS Publ*, **351**, 255–260.

Tawn, J., Shooter, R., Towe, R., & Lamb, R. 2018. Modelling spatial extreme events with environmental applications. *Spatial Statistics*, **28**, 39–58.

Thibaud, E., & Opitz, T. 2015. Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, **102**(4), 855–870.

Thibaud, E., Mutzner, R., & Davison, A.C. 2013. Threshold modeling of extreme spatial rainfall. *Water Resources Research*, **49**, 4633–4644.

Varin, C. 2008. On composite marginal likelihoods. *Advances in Statistical Analysis*, **92**(1), 1–28.

Varin, C., & Vidoni, P. 2005. A note on composite likelihood inference and model selection. *Biometrika*, **92**(3), 519–528.

Wadsworth, J.L., & Tawn, J.A. 2012. Dependence modelling for spatial extremes. *Biometrika*, **99**(2), 253–272.

Wadsworth, J.L., & Tawn, J.A. 2013. A new representation for multivariate tail probabilities. *Bernoulli*, **19**(5B), 2689–2714.

Wadsworth, J.L., & Tawn, J.A. 2019. *Higher-dimensional spatial extremes via single-site conditioning*. Preprint.

Wikle, C.K., Zammit-Mangion, A., & Cressie, N. 2019. *Spatio-Temporal Statistics with R*. Chapman & Hall/CRC The R Series. CRC Press.

Wolpert, R.L., & Ickstadt, K. 1998. Poisson/Gamma random fields for spatial statistics. *Biometrika*, **85**(2), 251–267.

Troisième partie

Articles annexés

Annexe A

G7 - A flexible model for spatial extremes. JSPI (2016).

Journal of Statistical Planning and Inference 172 (2016) 36–52

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

A flexible dependence model for spatial extremes

Jean-Noel Bacro [a,*], Carlo Gaetan [b], Gwladys Toulemonde [a]

[a] IMAG, Université de Montpellier, Montpellier, France
[b] DAIS, Università Ca' Foscari - Venezia, Italy

ARTICLE INFO

Article history:
Received 18 August 2014
Received in revised form 30 November 2015
Accepted 9 December 2015
Available online 24 December 2015

Keywords:
Spatial extremes
Asymptotic independence
Max-stable processes

ABSTRACT

Max-stable processes play a fundamental role in modeling the spatial dependence of extremes because they appear as a natural extension of multivariate extreme value distributions. In practice, a well-known restrictive assumption when using max-stable processes comes from the fact that the observed extremal dependence is assumed to be related to a particular max-stable dependence structure. As a consequence, the latter is imposed to all events which are more extreme than those that have been observed. Such an assumption is inappropriate in the case of asymptotic independence. Following recent advances in the literature, we exploit a max-mixture model to suggest a general spatial model which ensures extremal dependence at small distances, possible independence at large distances and asymptotic independence at intermediate distances. Parametric inference is carried out using a pairwise composite likelihood approach. Finally we apply our modeling framework to analyze daily precipitations over the East of Australia, using block maxima over the observation period and exceedances over a large threshold.

1. Introduction

The last decade has registered a considerable effort to model extremes of data collected through a collection of sites and the interested reader is referred to Bacro and Gaetan (2012) and Davison et al. (2012) for recent reviews.

If the main interest is producing return level maps, the modeling issue is mainly concentrated on relating the parameters of the marginal distributions in each site to geographical covariates. In case of a residual dependence, uncertainty of the estimates can be further adjusted (Fawcett and Walshaw, 2007). Additionally, this modeling approach can be extended to be hierarchical adding a layer for incorporating spatial dependence through a spatial random process (Casson and Coles, 1999; Cooley et al., 2007; Gaetan and Grigoletto, 2007; Sang and Gelfand, 2010).

If we are interested in modeling the joint occurrence of extremes over a region, then the dependence structure of a multivariate variable needs to be explicitly stated. In this case the usual modeling strategy consists in two steps (1) estimating the marginal distribution and (2) characterizing the dependence via a model issued by the multivariate extreme value (MEV) theory (see for example Beirlant et al., 2004, and the references therein). These two steps can be integrated in a proper inferential analysis (Padoan et al., 2010; Ribatet et al., 2012).

The MEV theory deals with the tail behavior of a multivariate distribution from which we pretend that a sample is drawn and distinguishes three different forms of extremal dependence: asymptotic dependence, asymptotic independence and exact independence.

* Corresponding author.
E-mail address: jean-noel.bacro@univ-montp2.fr (J.-N. Bacro).

http://dx.doi.org/10.1016/j.jspi.2015.12.002
0378-3758/

Asymptotic independence and dependence between a pair of random variables Z_1 and Z_2, with marginal distributions F_1 and F_2, can be defined in terms of

$$\chi = \lim_{u \to 1^-} \Pr(F_1(Z_1) > u | F_2(Z_2) > u), \tag{1}$$

where $\chi = 0$ and $\chi > 0$ represent asymptotic independence and dependence, respectively.

An example of a multivariate distribution which is asymptotically independent is given by the multivariate Gaussian distribution when the components are not perfectly correlated (Sibuya, 1960).

However the multivariate framework is inadequate for predicting or simulating values at unobserved sites. Therefore in the last years there was a general consensus in representing extreme spatial variability by max-stable processes (de Haan, 1984; Smith, 1990; Schlather, 2002; Kabluchko et al., 2009; Opitz, 2013) that are an infinite dimensional generalization of multivariate distributions for the maxima. The drawback of these processes is that they admit only two types of dependence structures in their finite dimensional distributions: asymptotic dependence or exact independence. This restriction is constraining when we model the tail behavior of the multivariate distribution of the data because it is difficult to assess in practice whether a data set should be modeled using an asymptotically dependent or asymptotically independent model (see Thibaud et al., 2013; Davison et al., 2013, for recent examples of these difficulties).

For coping with dependence structures that have not converged to their limiting form at observable levels, Wadsworth and Tawn (2012) introduced the hybrid spatial dependence model. The model originates from a max-mixture, namely $Z(s) = \max(\beta X(s), (1-\beta)Y(s))$ with $0 \leq \beta \leq 1$, of an asymptotically dependent process X (a max-stable process, for instance), and an asymptotically independent process Y.

In applications such as environmental ones different types of extremal dependencies could be present according the distance between two locations. As motivating example we shall consider winter daily cumulative rainfall, recorded at 31 monitoring sites in the East of Australia (Fig. 3). We quantify the strength of the asymptotic dependence of a pair of random variables $Z(s)$ and $Z(s+h)$, located at sites s and $s+h$, assuming the same marginal distribution F, by means of the dependence measure (Coles et al., 1999)

$$\chi(h,u) = 2 - \frac{\log \Pr(F(Z(s+h)) < u, F(Z(s)) < u)}{\log \Pr(F(Z(s)) < u)}, \quad 0 \leq u \leq 1. \tag{2}$$

In case of asymptotic independence, $\chi(u,h) \simeq 0$ for $u \simeq 1$ and $\chi(h,u)$ is zero for exactly independent variables for all u. Discriminating between asymptotic dependence and asymptotic independence by means of the estimates of $\chi(h,u)$ or $\chi(h)$, its limit version when $u \to 1^-$, is not easy, especially for rainfall extremes (Serinaldi et al., 2014). However the nonparametric estimates of $\chi(h,u)$ (see Fig. 5) suggest that asymptotic dependence is present up to a distance r_1 and asymptotic independence prevails for distances between r_1 and r_2 whereas for larger distances, exact independence could also be conjectured ($r_1 = 500$ km and $r_2 = 1000$ km, approximately, in Fig. 5).

In Wadsworth and Tawn (2012) the authors discuss the idea of having asymptotic dependence present up to a finite lag but in the reported examples asymptotic dependence or asymptotic independence are present for any distance, even if it is dimming with the distance. Following their idea, the contribution of the present paper is to consider examples of max-mixture between a max-stable process, that yields exact independence between maxima after a finite spatial lag and an asymptotically independent process that may or not yield exact independence between observations after that lag.

The max-stable process stems from the construction in Schlather (2002, p. 39) and, as example, we use the truncated Extremal Gaussian process (see also Davison and Gholamrezaee, 2012). For the asymptotically independent process we can consider stationary spatial processes with bivariate distributions satisfying only a general condition on the bivariate survivor functions (Ledford and Tawn, 1996, 1997). We exemplify our construction by means of a marginal transformed Gaussian process with possible finite range covariance function and of an inverse truncated Extremal Gaussian process.

The paper is organized as follows. In Section 2 we briefly introduce the max-stable and asymptotically independent processes and some classical extremal dependence measures. Our modeling proposal and its main properties are detailed in Section 3 and a pairwise likelihood approach is presented for the statistical inference. In Section 4 we show, by means of a simulation study, that this approach seems effective in order to identifying different max-mixture models. Section 5 is devoted to illustrate the modeling issues related to our motivating example. Concluding remarks and some perspectives are addressed in Section 6.

2. Spatial extremes modeling

2.1. Models for asymptotic dependence

Max-stable processes (de Haan, 1984) are an infinite-dimensional generalization of multivariate extreme value theory. The stochastic process $X = \{X(s), s \in \mathcal{D}\}$, where $\mathcal{D}$ is a spatial domain, is a max-stable process if and only if there exist functions $a_n(\cdot) > 0$ and $b_n(\cdot)$ on $\mathbb{R}$ such that

$$\max_{1 \leq i \leq n} \frac{X_i(s) - b_n(s)}{a_n(s)} \overset{d}{=} X(s)$$

where $X_1, X_2, \ldots$ are independent copies of X. In the sequel and without loss of generality, $\mathcal{D}$ is a subset of $\mathbb{R}^2$ and univariate margins of max-stable processes are assumed to be unit Fréchet, i.e. $\Pr(X(s) \leq x) = \exp(-1/x), x > 0$.

A max-stable process has a spectral representation (de Haan, 1984; Schlather, 2002). Assume that $r_i, i \geq 1$, are points of a Poisson process on $(0, \infty)$ with intensity dr. Let $W_i, i \geq 1$, be independent and identically distributed (i.i.d.) copies of a real valued continuous random function $W = \{W(s), s \in \mathcal{D}\}$, independent of the $\{r_i\}$ and such that $\mathbb{E}[W^+(s)] = \mu \in (0, \infty)$, where $W^+(s) = \max(W(s), 0)$. Then

$$X(s) = \mu^{-1} \max_{i \geq 1} W_i^+(s)/r_i \tag{3}$$

is a max-stable process with unit Fréchet margins.

Choosing a particular expression for W_i leads to known examples of max-stable processes: the Gaussian extreme value process (Smith, 1990), the extremal Gaussian process (Schlather, 2002), the Brown–Resnick process (Kabluchko et al., 2009) and the extremal t process (Opitz, 2013).

In the sequel we focus on a particular instance of a max-stable process: the Truncated Extremal Gaussian (TEG) process. The TEG process has been introduced by Schlather (2002) and has been exemplified by Davison and Gholamrezaee (2012). As in the extremal Gaussian model the model derives from an underlying Gaussian process censored on a compact random set.

Let $r_i, i \geq 1$, be defined as previously and consider $W_i(s) = c \max(0, \varepsilon_i(s)) \mathbb{I}_{B_i}(s - U_i)$ with ε_i independent copies of a stationary Gaussian process $\varepsilon = \{\varepsilon(s), s \in \mathcal{D}\}$ with zero mean, unit variance and correlation function $\rho(\cdot)$, $\mathbb{I}_B$ is the indicator function of a compact random set $B \subset \mathcal{D}$, of which B_i are independent replicates and U_i are points of a homogeneous Poisson process of unit rate on $\mathcal{D}$, independent of the ε_i. Choosing the constant c such that $c^{-1} = \mathbb{E}\left(\max\{W_i(s), 0\} \mathbb{I}_{B_i}(x - U_i)\right)$, the TEG process is defined as

$$X(s) = \max_{i \geq 1} \frac{W_i(s)}{r_i}. \tag{4}$$

The marginal distribution of X is unit Fréchet and the bivariate one is given by

$$\begin{aligned} &\Pr\left(X(s) \leq t_1, X(s+h) \leq t_2\right) \\ &= \exp\left\{-\left(\frac{1}{t_1} + \frac{1}{t_2}\right)\left[1 - \frac{\alpha(h)}{2}\left(1 - \left(1 - 2\frac{(\rho(h)+1)t_1 t_2}{(t_1+t_2)^2}\right)^{1/2}\right)\right]\right\} \end{aligned} \tag{5}$$

where $\alpha(h) = \mathbb{E}[|B \cap (h + B)|]/\mathbb{E}[|B|]$. If B is a disk of fixed radius r, $\alpha(h)$ can be approximated by

$$\alpha(h) \simeq (1 - \|h\|/(2r))\, \mathbb{I}_{[0,2r]}(\|h\|). \tag{6}$$

2.2. Models for asymptotic independence

A multivariate vector is asymptotically independent (AI) if and only if all its pairs of components are AI (de Oliveira, 1962). As a consequence, if all the bivariate distributions of a stochastic process are AI, the stochastic process is said to be AI.

For modeling AI we assume a specific model for bivariate joint tails as described in more detail in Ledford and Tawn (1996).

We assume that $\{Z(s), s \in \mathcal{D}\}$ is a stationary spatial process with unit Fréchet margins. Under weak conditions, Ledford and Tawn (1997, 1998) showed that the bivariate tail distribution of a pair of observations at s and $s + h$ admits an approximation such that

$$\Pr(Z(s) > z_1, Z(s+h) > z_2) \sim z_1^{-c_h^{(1)}} z_2^{-c_h^{(2)}} \mathcal{L}'_h(z_1, z_2)$$

for z_1 and z_2 simultaneously large, where $0 < 1/(c_h^{(1)} + c_h^{(2)}) \leq 1$ and $\mathcal{L}'_h \neq 0$ a bivariate slowly varying function (Bingham et al., 1987, Appendix 1) with limit function g_h such that for all $x > 0, y > 0$ and $c > 0, g_h(x, y) = \lim_{t \to \infty} \mathcal{L}'_h(tx, ty)/\mathcal{L}'_h(x, y)$ and $g_h(cx, cy) = g_h(x, y)$. The homogeneity property of g_h implies that $g_h(x, y) = g_h^*(w)$ with $w = x/(x+y) \in [0, 1]$ where the function g_h^*, often called the ray dependence function, is assumed to be a slowly varying function at 0 and 1.

Assuming $z_1 = z_2 = z$ leads to the Ledford–Tawn (LT) model for bivariate joint tails (Ledford and Tawn, 1996):

$$\Pr(Z(s) > z, Z(s+h) > z) \sim z^{-1/\eta(h)} \mathcal{L}_h(z) \quad \text{for } z \to \infty \tag{7}$$

where $\mathcal{L}_h(\cdot)$ is a univariate slowly varying function. The coefficient $\eta(h)$ varies between 0 and 1 and determines the decay rate of the bivariate tail probability $\Pr(Z(s) > z, Z(s+h) > z)$ for large z. Despite its simplicity, Eq. (7) appears as a very general model for bivariate joint tails which can provide, as detailed below, a measure of the extremal dependence between $Z(s)$ and $Z(s+h)$ through the coefficient $\eta(h)$. Asymptotic independence corresponds to $\eta(h) < 1$ and in such a case, $\eta(h)$ measures the degree of dependence in the asymptotic independence at h, where $\eta(h) > 1/2$ and $\eta(h) < 1/2$ indicate

positive and negative association, respectively. When the variables $Z(s)$ and $Z(s+h)$ are independent $\eta(h) = 1/2$. There are few examples of AI processes in the literature. In the sequel three asymptotically independent processes are considered with explicit expressions of (7).

Example 1: *Stationary Gaussian process*

Let $\{Y(s), s \in \mathcal{D}\}$ be a stationary Gaussian process with zero mean, unit variance and correlation function $\rho(h)$. Because bivariate Gaussian variables are AI provided that they are not perfectly correlated (Sibuya, 1960), the spatial process $Z'(s) = -1/\log(\Phi(Y(s)))$ has unit Fréchet margins and verifies (Ledford and Tawn, 1996)

$$\Pr(Z'(s) > z, Z'(s+h) > z) \sim C_h z^{-2/\{1+\rho(h)\}} (\log z)^{-\rho(h)/\{1+\rho(h)\}}$$

with $C_h = (1+\rho(h))^{3/2}(1-\rho(h))^{-1/2}(4\pi)^{-\rho(h)/\{1+\rho(h)\}}$. So $\eta(h) = \{1+\rho(h)\}/2$.

Example 2: *Inverse max-stable process*

The inverse max-stable process (Wadsworth and Tawn, 2012) is obtained by simply inverting a max-stable process. More precisely, let $\{X(s), s \in \mathcal{D}\}$ be a max-stable process defined as in (3). Then the process

$$Z(s) = -1/\log[1 - \exp\{-1/X(s)\}]$$

is an AI process with Fréchet margins. For any fixed h, the tail dependence coefficient is $\eta(h) = 1/\theta(h)$ where $\theta(h)$ is the extremal coefficient of the max-stable process.

Example 3: *Max-Gaussian ratio process*

Recently Padoan (2013) introduced a new family of spatial processes whose univariate limit distributions are unit Fréchet and bivariate distributions are able to cope with different levels of dependence according to the magnitude of extreme events. Such processes, called max-Gaussian ratio processes, are obtained as pointwise maxima of samples from a ratio of Gaussian processes with common correlation function. For every $n \in \mathbb{N}$, let $\{U_n(s), s \in \mathcal{D}\}$ and $\{V_n(s), s \in \mathcal{D}\}$ be two independent Gaussian processes on $\mathcal{D}$ with mean zero, unit variance and common correlation function, $\rho_n(h)$, such that

$$\rho_n(h) = 1 - \frac{\lambda(h)^2}{2n^2} + o\left(n^{-2}\right), \quad \text{as } n \to \infty.$$

Here $\lambda(h) > 0$ for $\|h\| \neq 0$. Assume also that $Y_{i,n}(s)$ are independent copies of $Y_n(s) = U_n(s)/V_n(s)$ and define $M_n(s) = \max_{i=1,\ldots,n} Y_{i,n}(s)$. Then the normalized bivariate asymptotic distribution of $(M_n(s), M_n(s+h))$ is

$$\begin{aligned}\Pr(W(s) \le w_1, W(s+h) \le w_2) &\equiv \lim_{n\to\infty} \Pr\left(M_n(s) \le \frac{nw_1}{\pi}, M_n(s+h) \le \frac{nw_2}{\pi}\right) \\ &= \exp\left\{-V_{\lambda(h)}(w_1, w_2)\right\}\end{aligned}$$

where

$$V_{\lambda(h)}(w_1, w_2) = \frac{1}{2}\left(\frac{2}{w_1} + \frac{2}{w_2} + \lambda(h) + \sqrt{\left(\frac{1}{w_1} - \frac{1}{w_2}\right)^2 + \lambda(h)^2} - \sqrt{\frac{1}{w_1^2} + \lambda(h)^2} - \sqrt{\frac{1}{w_2^2} + \lambda(h)^2}\right).$$

For a given $\lambda(h)$,

$$\Pr(W(s) > w, W(s+h) > w) \sim \left(1 + \frac{1}{2\lambda(h)}\right) w^{-2} \quad \text{as } w \to \infty$$

leading to a constant tail dependence parameter $\eta(h) = 1/2$. As underlined by Padoan (2013), a general framework based on Eq. (7) allows for different speeds of convergence to the independence case and could be used for dependence structures with a slower convergence than that of max-Gaussian ratio processes.

2.3. Pairwise extremal dependence measures

We recall here some measures of extremal dependence for spatial processes. From a theoretical point of view, the dependence structure of any multivariate extreme distribution is characterized by the exponent measure (see Resnick, 1987, for example). Unfortunately, it is quite difficult to infer this measure from the data. That is why summaries of extremal dependence based on pairwise measures have been proposed (Coles et al., 1999). For a stationary spatial process $Z = \{Z(s), s \in \mathcal{D}\}$ with univariate cumulative distribution function F, the pairwise extremal dependence between two sites s and $s+h$ can be characterized using the function

$$\chi(h) = \lim_{u\to 1^-} \Pr(F(Z(s+h)) > u \mid F(Z(s)) > u)$$

since we have pairwise asymptotic independence or asymptotic dependence (AD) if and only if $\chi(h) = 0$ or $\chi(h) \neq 0$, respectively (Sibuya, 1960). Alternatively $\chi(h)$ can be expressed as limit for $u \to 1^-$ of $\chi(h,u)$, defined in (2). The function $\chi(h,\cdot)$ can be interpreted as a quantile-dependent measure of dependence between two sites separated by h, giving more

insight if $Z(s)$ and $Z(s+h)$ are positively or negatively associated (Coles et al., 1999). Note also that for a max-stable process any bivariate distribution is max-stable and then the function $\chi(h,u)$ is constant with respect to u for a fixed h.

The extremal coefficient function (Schlather and Tawn, 2003) is a specific measure of the dependence for a max-stable process X. Given a pair of sites s and $s+h$ the extremal coefficient function $\theta(h)$ is defined as

$$\Pr(X(s) \le x, X(s+h) \le x) = \Pr(X(s) \le x)^{\theta(h)}.$$

Here $1 \le \theta(h) \le 2$ and $\theta(h) = 1$ or $\theta(h) = 2$ corresponds to perfect dependence or exact independence, respectively. It is easy to show that $\theta(h) = 2 - \chi(h)$.

Special instances of the Gaussian extreme value process (Smith, 1990) or the Brown–Resnick process (Kabluchko et al., 2009) span the range of possible extremal dependencies from perfect dependence to exact independence provided that distance $\|h\|$ increases indefinitely. Instead the extremal Gaussian process (Schlather, 2002) cannot account for extremes that become independent after some distance. Note that the TEG process has the feature that its extremal coefficient function

$$\theta(h) = 2 - \alpha(h)\left\{1 - 2^{-1/2}[1-\rho(h)]^{1/2}\right\} \tag{8}$$

reaches the upper value ($\theta(h) = 2$) for $\|h\|$ large enough. This specific feature will be exploited later.

Under asymptotic independence, both $\chi(h)$ and $\theta(h)$ functions are uninformative and of limited interest. Assume again that Z is a stationary spatial process with univariate cumulative distribution function F, and define

$$\bar{\chi}(h,u) = \frac{2\log\Pr(F(Z(s)) > u)}{\log\Pr(F(Z(s)) > u, F(Z(s+h)) > u)} - 1, \quad 0 \le u \le 1. \tag{9}$$

The limit $\bar{\chi}(h) = \lim_{u\to 1^-}\bar{\chi}(h,u)$, with $-1 < \bar{\chi}(h) \le 1$, provides another measure that increases with the extremal dependence between $Z(s)$ and $Z(s+h)$ (Coles et al., 1999). It turns out that for AD process $\bar{\chi}(h) = 1$, for all h. Moreover under the condition (7), it is easy to show that $\bar{\chi}(h) = 2\eta(h) - 1$ and the tail dependence coefficient $\eta(h)$ appears as another dependence measure of interest (Ledford and Tawn, 1996, 1997; Ancona-Navarrete and Tawn, 2002).

Note that the empirical estimate of (9) provides a useful statistic for inspecting the tail behavior when $u < 1$. For the stationary Gaussian process with correlation function $\rho(h)$ we can show that $\bar{\chi}(h,u)$ varies with u (Coles et al., 1999, p. 348) with limit $\bar{\chi}(h) = \rho(h)$ and $\eta(h) = (1+\rho(h))/2$. For the inverse max-stable process, $\bar{\chi}(h,\cdot)$ is a constant function. In other words, bivariate survival distributions of inverse max-stable processes are uniquely linked to the marginal survival function of the process whatever the magnitude of the considered extreme events. Moreover we have $\bar{\chi}(h) = 2/\theta(h) - 1$.

Finally, the function $\bar{\chi}(h,u)$ of a max-Gaussian ratio process varies with u and tends to 0 as $u \to 1^-$ for a fixed value of $\lambda(h)$.

3. Max-mixture modeling of spatial extremal dependence

3.1. Model specification

Let $X = \{X(s), s \in \mathcal{D}\}$ and $Y = \{Y(s), s \in \mathcal{D}\}$ be two independent stationary spatial processes, such that X is a max-stable process and Y an AI process both with unit Fréchet univariate distributions. We define the max-mixture (MM) model as

$$Z(s) = \max(\beta X(s), (1-\beta)Y(s)), \quad 0 \le \beta \le 1. \tag{10}$$

The MM model has been introduced by Wadsworth and Tawn (2012) for modeling situations where the extremal dependence structure may vary with distance. Even if it is not max-stable process, the MM model allows a different order of decay towards an asymptotically dependent limit which inherits the same dependence structure of X. In Wadsworth and Tawn (2012) various instances of max-stable processes along with their inverted versions as AI processes have been considered and all fitted models had asymptotic dependence or asymptotic independence present at all spatial lags.

Owing to our motivating data set, we propose in the sequel to extend the set of examples by considering a max-mixture model that deals with asymptotic dependence at short lags, asymptotic independence at intermediate lags and possibly exact independence at larger lags. More precisely we choose as X a TEG process (4) with covariance function $\rho(\cdot)$. Moreover, with respect to the examples in Wadsworth and Tawn (2012), we broaden the class of considered AI processes by taking into account AI processes with unit Fréchet univariate distributions and bivariate distributions satisfying the LT model (7) for $\eta(h) < 1$.

Using the independence between the two processes X and Y it is straightforward to obtain the bivariate distribution for a pair of sites, namely

$$\begin{aligned}&\Pr(Z(s) \le z_1, Z(s+h) \le z_2)\\ &= \exp\left\{-\beta\left(\frac{1}{z_1}+\frac{1}{z_2}\right)\left[1 - \frac{\alpha(h)}{2}\left(1 - \left(1 - 2\frac{(\rho(h)+1)z_1z_2}{(z_1+z_2)^2}\right)^{1/2}\right)\right]\right\} F_Y^h\left(\frac{z_1}{1-\beta}, \frac{z_2}{1-\beta}\right)\end{aligned} \tag{11}$$

where $F_Y^h(y_1,y_2) = \Pr(Y(s) \le y_1, Y(s+h) \le y_2)$. Since $\Pr(Z(s) \le z) = \Pr(Z(s) \le z, Z(s+h) < \infty) = \exp(-1/z)$ the model has unit Fréchet univariate distribution.

3.2. Pairwise extremal dependence measures of the model

Exploiting characterization (7), the bivariate tail distribution of (10), for large z, can be expressed as:

$$\Pr(Z(s) > z, Z(s+h) > z) = \frac{\beta(2-\theta(h))}{z} + \left(\frac{z}{1-\beta}\right)^{-1/\eta(h)} \mathcal{L}_h\left(\frac{z}{1-\beta}\right) + O(z^{-2}).$$

So it is easy to deduce the $\chi(h)$ function using Eq. (8), namely

$$\chi(h) = \beta\,(2-\theta(h)) = \beta\,\alpha(h)\left(1-\sqrt{\frac{1-\rho(h)}{2}}\right).$$

If the approximation (6) holds, it turns out that pairs of sites separated by a distance $\|h\|$ are AD if this distance is smaller than $2r$ and AI otherwise.

For evaluating $\bar{\chi}(h)$, we need to evaluate the logarithm of the bivariate tail distribution. We obtain

$$\log \Pr(Z(s) > z, Z(s+h) > z) = \begin{cases} \log\left(\beta(2-\theta(h))\right) - \log z + o(\log(z)) & \text{if } 2-\theta(h) \neq 0 \\ -\eta(h)^{-1}\log\left(\frac{z}{1-\beta}\right) + \log \mathcal{L}_h\left(\frac{z}{1-\beta}\right) + o(1), & \text{otherwise.} \end{cases}$$

If $2-\theta(h) \neq 0$, we can conclude that $\bar{\chi}(h,z) \to 1$ as $z \to \infty$. On the other hand, if $2-\theta(h) = 0$, we have

$$\bar{\chi}(h,z) \sim \frac{\left(-2-\frac{2}{z\log z}\right)}{\left(-\frac{1}{\eta(h)}\left(1-\frac{\log(1-\beta)}{\log z}\right)+\frac{\log(\mathcal{L}_h(z/(1-\beta)))}{\log z}\right)} - 1,$$

i.e. $\bar{\chi}(h,z) \to 2\eta(h) - 1$ as $z \to \infty$. Owing to (6) the results can be summarized into the formula

$$\bar{\chi}(h) = \mathbb{I}_{[0,2r]}(\|h\|) + (2\eta(h)-1)\mathbb{I}_{[2r,\infty)}(\|h\|),$$

that highlights the different behavior according to the distance between two sites. Let $R > 2r$ and assume that $\eta(h) = 1/2$ for $\|h\| > R$. Then pairs of sites separated by a distance $\|h\|$ are asymptotically dependent for $\|h\| < 2r$, asymptotically independent for $2r \leq \|h\| \leq R$ and near independent for $\|h\| > R$. For example, for the transformed stationary Gaussian process with unit Fréchet margins and correlation function $\rho_Y(h)$, we have:

$$\bar{\chi}(h) = \mathbb{I}_{[0,2r]}(\|h\|) + \rho_Y(h)\mathbb{I}_{[2r,\infty)}(\|h\|).$$

In that case, independence is achieved if the correlation function $\rho_Y(\cdot)$ is such that $\rho_Y(h) = 0$ when $\|h\| > R$.

3.3. Model inference

For the model (10) since the full likelihood is intractable to evaluate, a composite likelihood approach is used for parametric estimations using pairs. The composite likelihood is an inference function derived by multiplying likelihoods of marginal or conditional events (Lindsay, 1988; Varin, 2008). Such an approach has been applied in spatial extremes using bivariate densities of max-stable processes (Padoan et al., 2010) or bivariate density of exceedances over a large threshold (Jeon and Smith, 2012; Wadsworth and Tawn, 2012; Bacro and Gaetan, 2014; Huser and Davison, 2014). Recently, improvements have been obtained for the parameters estimations of some max-stable processes, e.g. Brown–Resnick processes: extremal increments of the process allow to work with a complete likelihood function (Engelke et al., 2015; Wadsworth and Tawn, 2014). A direct modeling of the exceedances of a max-stable process is also possible using a generalized Pareto process (Ferreira and de Haan, 2014) but such an approach is only of interest in the case of asymptotic dependence.

If z_{ik} is the site-wise block maximum, for instance seasonal maximum, observed at site s_i, $i = 1,\ldots,N$ and at time t_k, $k = 1,\ldots,M$, the pairwise (weighted) log-likelihood is defined by

$$\mathrm{pl}(\psi) = \sum_{k=1}^{M} \mathrm{pl}_k(\psi) = \sum_{k=1}^{M}\sum_{i=1}^{N-1}\sum_{j>i}^{N} w_{ij}\log L(z_{ik}, z_{jk}; \psi) \tag{12}$$

where $L(z_{ik}, z_{jk}; \psi)$ is the likelihood of a pair z_{ik}, z_{jk}, The weights w_{ij} are non negative and specify the contributions of each pairs. A simple weighting choice is to let $w_{ij} = 0$ for any pair whose distance exceeds a specified value δ, and let $w_{ij} = 1$, otherwise.

Recently Wadsworth and Tawn (2012) argued that, under asymptotic independence, it is more natural to model the original events provided that they exceed a large threshold, u. Following their proposal the pairwise likelihood contribution $L(z_{ik}, z_{jk}; \psi)$ becomes

$$L(z_{ik}, z_{jk}; \psi) = \begin{cases} \dfrac{\partial^2}{\partial z_{ik} \partial z_{jk}} G(z_{ik}, z_{jk}; \psi) & \text{if } \max(z_{ik}, z_{jk}) > u \\ G(u, u; \psi) & \text{if } \max(z_{ik}, z_{jk}) \leq u \end{cases} \tag{13}$$

where z_{ik} is the observed value and $G(\cdot, \cdot)$ designates the bivariate distribution (11).

When dealing with exceedances it is not reasonable to assume that the observations are independent over the time. Assuming that the space–time process is temporally α mixing, the function (12) is a contrast function and conditions in Guyon (1995, Theorem 3.4.7) are satisfied. Thus the maximum composite likelihood estimator $\widehat{\psi}$ is asymptotically Gaussian for large M and its asymptotic variance is given by the inverse of the Godambe information matrix $\mathcal{G}(\psi) = \mathcal{H}(\psi)[\mathcal{J}(\psi)]^{-1}\mathcal{H}(\psi)$. Standard error evaluation requires consistent estimation of the matrices $\mathcal{H}(\psi) = \mathbb{E}(-\nabla^2 \mathrm{pl}(\psi))$ and $\mathcal{J}(\psi) = \mathbb{V}ar(\nabla \mathrm{pl}(\psi))$.

It is worth noting that such results hold if the data are actually from the limit model and this fact can add a bias (for an accurate study see Jeon and Smith, 2012) and, consequently, further uncertainty in the estimates.

The matrix $\mathcal{H}(\psi)$ can be estimated by $\widehat{\mathcal{H}} = -\nabla^2 \mathrm{pl}(\widehat{\psi}_T)$. Estimation of the matrix

$$\mathcal{J}(\psi) = \sum_{k=1}^{M} \sum_{k'>k}^{M} \mathbb{C}ov \left\{ \nabla \mathrm{pl}_k(\psi) \nabla \mathrm{pl}_{k'}(\psi)' \right\}$$

requires some care when we deal with temporally dependent data. In this paper we estimate $\mathcal{J}$ by means of a subsampling technique (Carlstein, 1986). More precisely, we consider B overlapping blocks $D_b \subset \{1, \ldots, M\}$, $b = 1, \ldots, B$, of size d_b and the estimate

$$\widehat{\mathcal{J}} = \frac{M}{B} \sum_{b=1}^{B} \frac{1}{d_b} \nabla \mathrm{pl}_{D_b}(\widehat{\psi}) \nabla \mathrm{pl}_{D_b}(\widehat{\psi})'$$

where pl_{D_b} is the pairwise likelihood evaluated over the block D_b.

Finally we mention that an appropriate model selection criterion to the pairwise likelihood is the composite likelihood information criterion (Varin and Vidoni, 2005), namely

$$\mathrm{CLIC} = -2 \left[\mathrm{pl}(\hat{\psi}) - \mathrm{tr}\{\hat{\mathcal{H}}^{-1} \hat{\mathcal{J}}\} \right].$$

Lower values of CLIC indicate better fit.

4. Simulation study

To assess the quality of the estimation procedure in case of the MM model (10), a simulation study has been carried out. We have chosen for X a TEG process (4) where B is a disk with a fixed radius r and exponential correlation function $\rho(h) = \exp(-\|h\|/\rho_1)$, $\rho_1 > 0$. The asymptotically independent process, Y, is given by $Y(s) = -1/\log(\Phi(Y'(s)))$, where Φ is the cumulative distribution function of a standard normal distribution and $\{Y'(s), s \in \mathcal{D}\}$ is a Gaussian spatial process with spherical correlation function, i.e. $1 - 1.5(\|h\|/\rho_2) + 0.5(\|h\|/\rho_2)^3$, for $\|h\| \leq \rho_2$, zero otherwise, $\rho_2 > 0$.

Under this setup extreme observations at sites separated by a vector h are asymptotically dependent if $\|h\| < r$, asymptotically independent if $r \leq \|h\| < \rho_2$ and independent if $\|h\| \geq \rho_2$, provided that $r < \rho_2$.

Five simulated images of the MM model over the $[0, 1]^2$ square are shown in Fig. 1, according to different values of the mixing parameter β. Actually, in order to appreciate the role of the mixing parameter β, the values in the images are derived by considering the simulation when $\beta = 0$ (AI process) and $\beta = 1$ (AD process). Note that the degree of smoothness decreases with β.

In the simulation study we have considered a moderately sized data set with $N = 49$ sites and $M = 1000$ independent observations. To avoid too systematic distances between pairs of sites, a non regular spatial grid has been considered. The $[0, 1]^2$ square is divided into 49 equal sub-squares and within each sub-square a point is uniformly chosen at random. We set $\rho_1 = 0.2$, $\rho_2 = 0.8$ and $r = 0.25$ and different values of $\beta \in \{0, 0.25, 0.50, 0.75, 1\}$.

The parameters are estimated on 500 data replication using the composite likelihood approach detailed in Section 3.3. The threshold u is taken corresponding to the 0.9 empirical quantile at each site and the δ value is chosen as the 0.9 quantile of the distribution of the distances between pairs of sites.

For compactness we report only the results for 500 data replications with $\beta = 0, 0.25, 0.75, 1$. For $\beta = 0.5$ we have obtained similar results. The boxplots in Fig. 2 that, overall, the parameters are well estimated except ρ_2 the parameter of the spherical correlation for which the estimate is significantly biased. This inadequacy seems consistent with the difficulties in estimating the parameter of the spherical correlation function in Gaussian models (Mardia and Watkins, 1989). In our example a justification for choosing a spherical correlation function is to consider a potential extremal exact independence

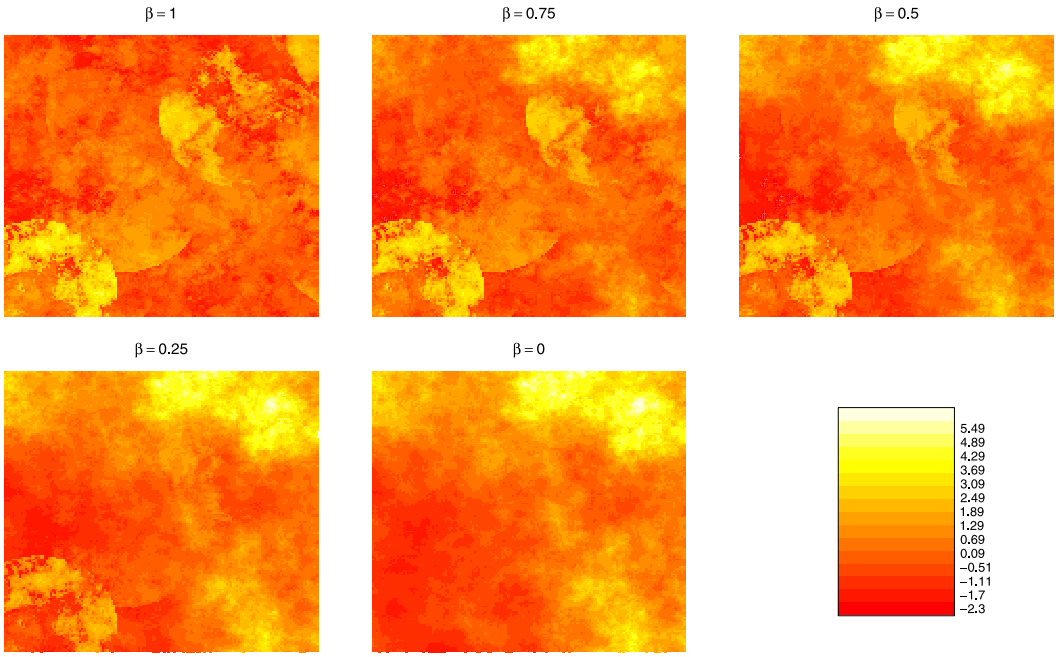


Fig. 1. Simulations of the MM_β model (10) on the logarithm scale according different values of $\beta \in \{0, 0.25, 0.50, 0.75, 1\}$. The compact set B is taken as a disk with a fixed radius $r = 0.25$. An exponential correlation function with parameter $\rho_1 = 0.2$ is chosen for the underlying Gaussian process. For the AI process a Gaussian random field is considered with a spherical correlation function with parameter $\rho_2 = 0.8$.

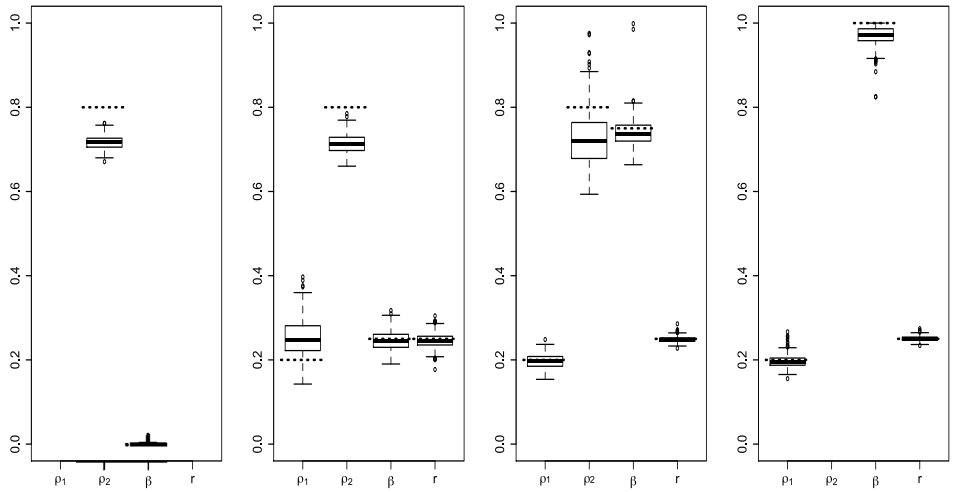


Fig. 2. Boxplots of 500 estimates from 1000 independent copies of the MM_β model (from left to right: $\beta = 0$, $\beta = 0.25$, $\beta = 0.75$ and $\beta = 1$) with $\rho_1 = 0.2$, $\rho_2 = 0.8$ and $r = 0.25$. For $\beta = 0$ and $\beta = 1$, only the results for the identifiable parameters are reported.

for distances larger than ρ_2. Simulations with an exponential correlation function not reported here lead to unbiased estimates of the range parameter.

Thereafter, we assessed whether CLIC is useful in identifying the true model, i.e. in our framework if we can use CLIC for discriminating between asymptotic independence, asymptotic dependence or a mixture of this. We have considered 500 simulations from mixture models with the same setting as before. In Table 1 we summarize our findings that are quite encouraging. We denote by MM_β, $\beta \in \{0, 0.25, 0.50, 0.75, 1\}$ the MM model according to different values of the mixing parameter. When the simulations come from MM_β, $\beta = 0.25, 0.5$ and 0.75, identification based on minimizing the CLIC value performs extremely well. Moreover the proportion of simulations in which the true model is detected is 68.6% if the true model is the AI process (MM_0). This proportion increases to 80% when the TEG process ($\beta = 1$) is the true model.

Table 1
Number of identified models according CLIC under different MM_β model, $\beta \in \{0, 0.25, 0.50, 0.75, 1\}$ with $\rho_1 = 0.2$, $\rho_2 = 0.8$, $r = 0.25$.

	Gaussian	MM	TEG
MM_0	346	154	0
$MM_{0.25}$	0	500	0
$MM_{0.50}$	0	500	0
$MM_{0.75}$	0	498	2
MM_1	0	100	400

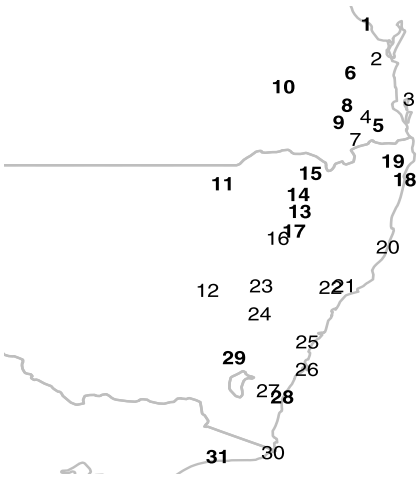


Fig. 3. Geographical locations of the 31 meteorological stations in the East Australia. Stations with a label in bold character are used for model inference and the other stations are put aside for validating the models.

5. Real data example

We analyze daily rainfall data from the 31 stations in the East of Australia whose locations are shown in Fig. 3. The values come from the daily rainfall data set of Lavery et al. (1992), available at time of writing at ftp.bom.gov.au/anon/home/ncc/www/change/HQdailyR.

Daily rainfall totals are for the 24-hours (measured at 9 am) and we consider days in the winter period (April–September) for 49 years ranging from 1955 to 2003.

Empirical estimates of the functions $\chi(h, u)$ and $\bar{\chi}(h, u)$ can be constructed on the basis of observed data by using the empirical estimates of univariate and bivariate distributions. In order to explore possible anisotropy of the dependence we have plotted (Fig. 4) the loess smoothing of $\hat{\chi}(h, u)$ and $\hat{\bar{\chi}}(h, u)$ at $u = 0.975$ with respect to the distances in different directional sectors, namely $(-\pi/8, \pi/8]$, $(\pi/8, 3\pi/8]$, $(3\pi/8, 5\pi/8]$, $(5\pi/8, 7\pi/8]$, where 0 represents the northing direction. Based on these estimates there is no clear evidence of anisotropy even if a stronger spatial dependence appears in the northing direction.

Moreover, as we mentioned in the introduction, the isotropic estimates (Fig. 5) of the functions $\hat{\chi}(h, u)$ and $\hat{\bar{\chi}}(h, u)$ at different values of the threshold u suggest that asymptotic dependence between sites seems to be present up to a distance of 500 km, and asymptotic independence could be conjectured between 500 and 1000 km distances. Therefore a max-mixture model seems a good candidate for interpreting the extreme value dependence. However the strength of dependence decreases when considering exceedances of increasing thresholds. This fact highlights the difficulty in a proper modeling of the asymptotic dependence for short distances.

In the sequel, we shall consider seven models that belong to three classes: max-mixture, max-stable and asymptotically independent. Each model is fitted using a subset of 16 sites and the remaining sites are used to perform model validation. We shall consider three MM models, namely

A_1 a MM model (10) specification in which X is a TEG process with exponential correlation function $\exp\{-\|h\|/\rho_1\}$, $\rho_1 > 0$ and B_1 is a disk of fixed and unknown radius r_1. The asymptotically independent process is given by $Y(s) = -1/\log((\Phi(Y'(s))))$, where Φ is the cumulative distribution function of a normalized Gaussian random variable and Y' is a Gaussian spatial process with spherical correlation function $1 - 1.5(\|h\|/\rho_2) + 0.5(\|h\|/\rho_2)^3$, for $\|h\| \leq \rho_2$, zero otherwise, $\rho_2 > 0$;

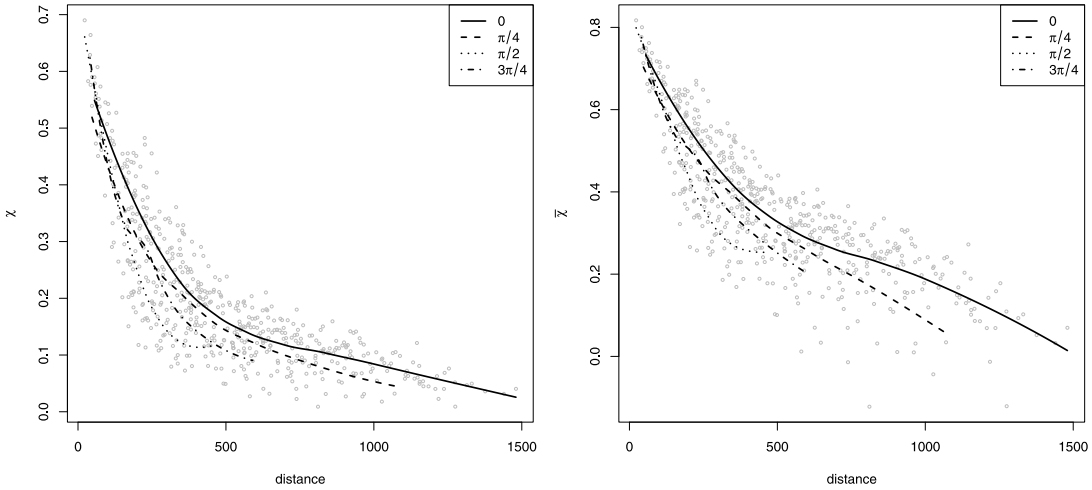


Fig. 4. Empirical evaluation of the functions $\widehat{\chi}(h,u)$ (left) and $\widehat{\bar{\chi}}(h,u)$ (right) at $u = 0.975$. Gray circles give empirical value between all available pairs. Lines represent smoothed values of the empirical estimates using the pairs in the directional sectors $(-\pi/8, \pi/8]$, $(\pi/8, 3\pi/8]$, $(3\pi/8, 5\pi/8]$ and $(5\pi/8, 7\pi/8]$.

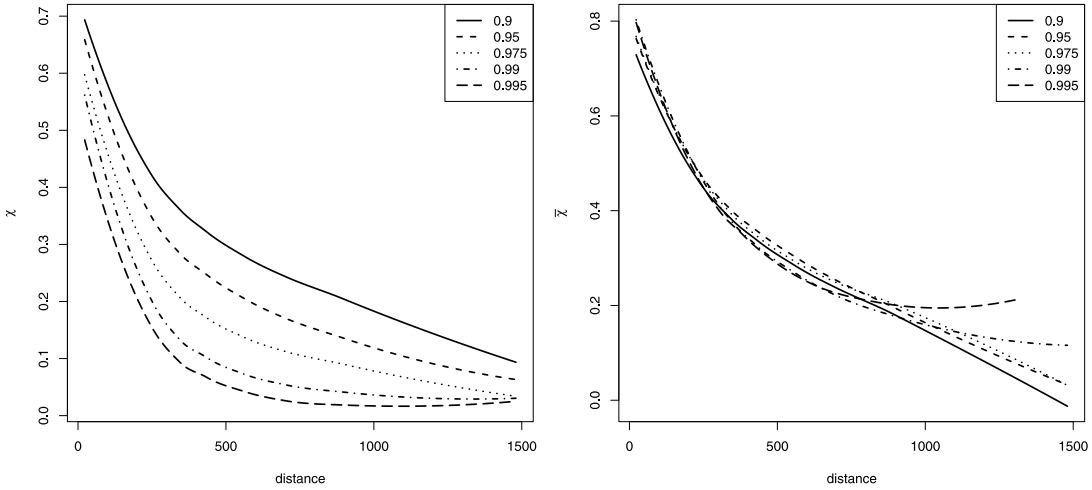


Fig. 5. Smoothed values of the empirical estimates of the functions $\hat{\chi}(h,u)$ (left) and $\hat{\bar{\chi}}(h,u)$ (right) at different values of the threshold u.

A_2 a MM model (10) where X is a TEG process as in A_1 and Y' is a Gaussian spatial process with exponential correlation function $\exp\{-\|h\|/\rho_2\}$;

A_3 a MM model with the same X as specified in A_1 and A_2 and in which Y is an inverse TEG process with exponential correlation function $\exp\{-\|h\|/\rho_2\}$, $\rho_2 > 0$. The B_2 disk has a fixed and unknown radius r_2.

As max-stable model candidate, we consider a max-stable model that entails exact independence between sites after a distance greater than $2r_1$, i.e.

B the TEG process specified in A_1.

Finally we take into account three asymptotically independent models, namely

C_1 the asymptotically independent process specified as Y in A_1;

C_2 the asymptotically independent process specified as Y in A_2;

C_3 the asymptotically independent process specified as Y in A_3.

Note that models C_1 and C_3 result in exact independence after distances greater than ρ_2 and r_2, respectively.

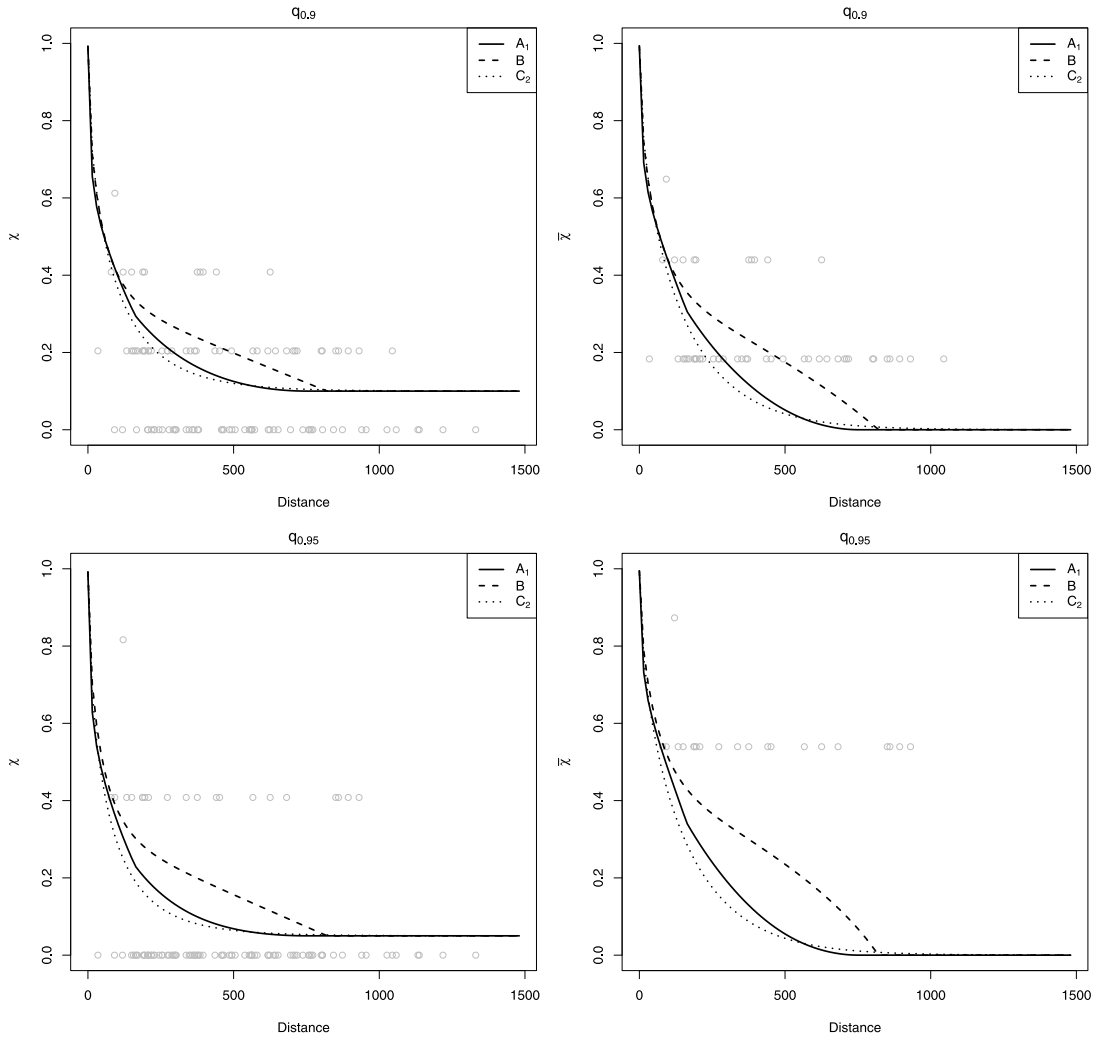


Fig. 6. Site-wise winter maxima: empirical and fitted values for $\hat{\chi}(h, u)$ and $\hat{\bar{\chi}}(h, u)$. Empirical values are computed using the validation data set. Top row: $u = 0.9$; bottom row: $u = 0.95$.

5.1. Site-wise maxima

First of all we have considered model site-wise winter maxima. Model (10) assumes common marginal Fréchet distributions and a proper inferential approach requires to fit marginal and dependence parameters. However, because we are interested in the appropriateness of different degrees of spatial asymptotic dependence, we prefer to follow a more simple and pragmatic approach. Specifically, we fit separately a GEV distribution in each site and we use the estimates to transform the marginals to unit Fréchet. The dependence parameters are estimated using the pairwise likelihood approach. Padoan et al. (2010) found in their simulation study that relatively small values of the distance δ in (12) produce gains in computation efficiency as well as in statistical efficiency of the estimates. However in our case we prefer to set $\delta \simeq 1000$ km, which entails to consider about 90% of all distinct observational pairs. For evaluating the CLIC and the standard errors we assume that the seasonal maxima are independent. In that case the estimation of the matrix $\mathcal{J}$ is greatly simplified in the subsampling procedure and we have considered $M = 49$ non overlapping blocks D_b corresponding to a single year, i.e. $d_b = 1$.

Our findings are summarized in Table 2. The rather wide standard-error of the spatial parameters, in particular for the max-mixture models, probably can be justified by the small number of independent replications over the years, pointing out that it is hard to separate the contribution of the components in the max-mixture. As suggested by the CLIC, the best-fitting model is A_1, for which pairs of sites separated by a distance d smaller than 160 km or greater than 750 km are

Table 2
Summary of the fitted models based on the site-wise winter maxima from the Australian data. Standard errors are reported between parentheses.

Model	$\widehat{\rho}_1$	$\widehat{r}_1$	$\widehat{\rho}_2$	$\widehat{r}_2$	$\widehat{\beta}$	CLIC
A_1	10.82 (14.10)	81.22 (301.93)	752.42 (278.52)	–	0.38 (0.22)	22 623.54
A_2	29.05 (37.56)	177.91 (32.16)	1451.92 (187.49)	–	0.72 (0.04)	22 661.76
A_3	5.47 (5.63)	311.22 (81.88)	367.48 (129.36)	707.73 (217.12)	0.43 (0.07)	22 644.5
B	78.09 (18.32)	410.34 (86.77)	–	–	–	22 692.3
C_1	–	–	359.51 (42.96)	–	–	22 689.73
C_2	–	–	179.34 (21.59)	–	–	22 642.23
C_3	–	–	71.84 (19.11)	440.48 (63.99)	–	22 679.72

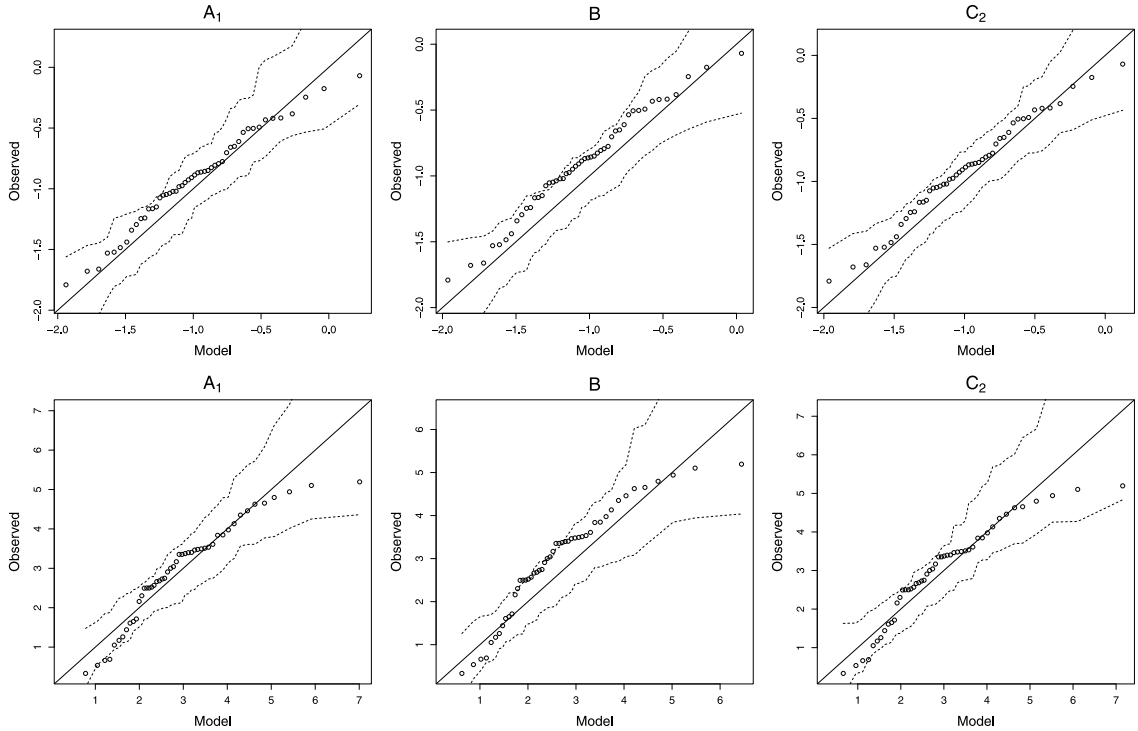


Fig. 7. Site-wise winter maxima: quantile–quantile plots for the minimum and maximum values on the validation data set (15 sites). The three columns correspond to fitted models A_1, B and C_2, respectively. The top row compares the minimum of the validation data set with its corresponding value under the fitted models. The bottom row compares in the same way the maximum values on the validation data set.

asymptotically max-stable dependent or exactly independent, respectively. At intermediate distances the seasonal maxima exhibit asymptotic independence. Moreover according to the CLIC values the MM models and the asymptotic independence models appear superior to the max-stable model B. So the max-stable model seems to overestimate the level of dependence in the data.

The goodness of fit has been also assessed in two different ways. Fig. 6 shows the empirical values for $\chi(h, u)$ and $\bar{\chi}(h, u)$, with $u = 0.9$ and 0.95 and their model-based counterparts of the three best models in each class according to the CLIC. Empirical estimates are calculated on the validation data set. The fits at finite thresholds are similar for A_1 and C_2 and the max stable model B entails stronger dependence for any distance. Considering the general patterns and owing to the small

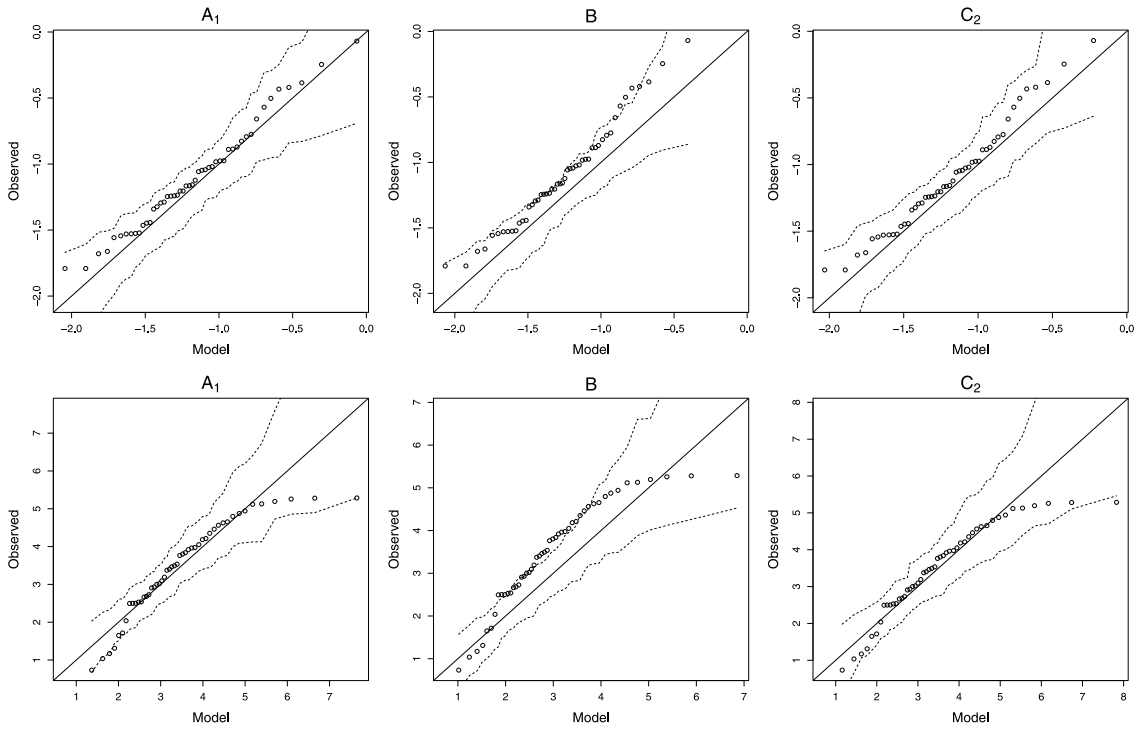


Fig. 8. Site-wise winter maxima: quantile–quantile plots for the minimum and maximum values on the complete data set (31 sites). The three columns correspond to fitted models A_1, B and C_2, respectively. The top row compares the minimum of the validation data set with its corresponding value under the fitted models. The bottom row compares in the same way the maximum values on the validation data set.

number of repeated observations for each site it is difficult to see which model catches better the bulk of the empirical values.

Model checking is also performed through QQ-plots for the logarithm of different groupwise minima and maxima on the validation set (Fig. 7) and the complete data (Fig. 8). Such plots provide some insight into whether the dependence models inferred using pairwise likelihood are capturing higher order dependence structures (Wadsworth and Tawn, 2012). Inspecting these plots, it appears that the multivariate distribution of the seasonal maxima is poorly modeled by the max-stable model B. Instead models A_1 and C_2 lead to quite similar results with an overall agreement to the data. Considering these plots and the CLIC values there is an overall evidence in favor of the max-mixture model.

5.2. Threshold exceedances

Now we deal with daily precipitations in the winter period and we use exceedances above a threshold corresponding to the 0.975 quantile in the empirical distribution at each site. We transform the observations to a unit Fréchet variable using the empirical distribution for data below the threshold and a site-wise fitted Generalized Pareto Distribution for data above the threshold. We estimate the spatial dependence parameters using the pairwise likelihood contribution (13). Because the original event data appear temporal dependent, the estimates $\hat{\mathcal{H}}$ and $\hat{\mathcal{J}}$ are carried out using a sliding temporal window of $d_b = 30$ days.

According to the CLIC value (Table 3), the preferred model is the *MM* model A_3. Nevertheless, the results for this model, here reported for completeness, have to be carefully considered because the estimate of β is virtually indistinguishable from zero, pointing out there is no mixture between the max-stable process and the asymptotically independent one. For $\beta = 0$ the parameters of the max-stable component are not identifiable and this fact affects the values of the estimates, their standard errors and finally the CLIC value. Moreover model A_3 reduces to model C_3 for $\beta = 0$ which corresponds the second best CLIC value, indicating some evidence for asymptotic independence for all distances.

Setting aside A_3 we reconsider the empirical and fitted values for $\hat{\chi}(h, u)$ and $\hat{\bar{\chi}}(h, u)$, $u = 0.9$ and $u = 0.95$, for the three best models in each class, namely A_2, B and C_3 (see Fig. 9). Model B seems to overestimate the asymptotic dependence at large distances. Again the fits of A_2 and C_3 look overall similar with a slight preference for A_2.

Finally, in order to illustrate the behavior of the models and check the fitting, we consider empirical and simulation based model estimates of few conditional probabilities. We choose the site s_1 in the top-right corner of the map (see

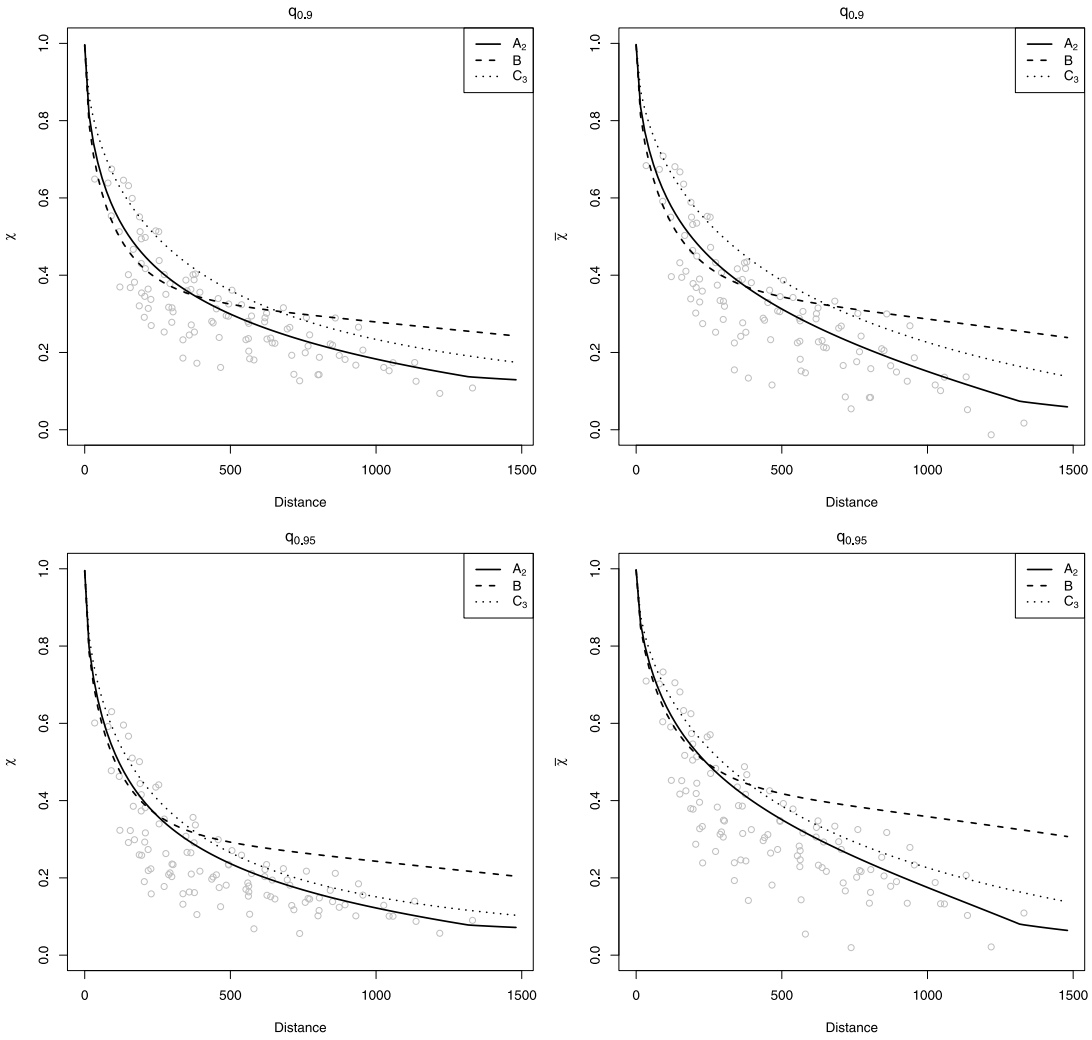


Fig. 9. Winter daily data: empirical and fitted values for $\hat{\chi}(h, u)$ and $\hat{\bar{\chi}}(h, u)$. Empirical values are computed using the validation data set and models are fitted using the q_u quantile exceedances. Top row: $u = 0.9$; bottom row: $u = 0.95$.

Fig. 3) as a reference location and we consider three subsets of sites $\mathcal{S}_1 = \{s_2, s_3, s_6, s_8, s_{10}\}$, $\mathcal{S}_2 = \{s_{11}, s_{13}, s_{14}, s_{15}, s_{18}\}$ and $\mathcal{S}_3 = \{s_{25}, s_{26}, s_{27}, s_{28}, s_{29}\}$ that roughly correspond to three different classes of distances from s_1. Then we compute the conditional probabilities $\Pr(Z(s) > z, s \in \mathcal{S}_i \mid Z(s_1) > z)$, $i = 1, 2, 3$ for different large values of p such that $\Pr(Z(s_1) \leq z) = p$. The confidence intervals in Fig. 10 are based on block bootstrapping of simulated daily data. The overall impression is that the max-stable model B is not able to describe the extremal dependence at medium ($\mathcal{S}_2$) and large distances ($\mathcal{S}_3$). Model C_3 basically overestimates the empirical probabilities for different thresholds and exhibits a lack of fitting for relative small distances ($\mathcal{S}_1$). On the other hand the fitting of model A_2 is more consistent at different thresholds and distances. Lastly note that both models agree for very large thresholds. These findings indicate that the max-mixture models we propose add modeling flexibility to spatial extreme analysis and seem able to encompass different degrees of spatial extremal dependence.

6. Conclusion

In this paper we have proposed a unifying spatial model which combines different degrees of extremal dependence depending on the distance between pairs of sites. Our approach exploits the max-mixture model proposed by Wadsworth and Tawn (2012) and focuses on the possible detection of pairwise max-stable dependence at short distances, asymptotic independence at intermediate ones and possibly exact independence at large distances. At short distances the extremal

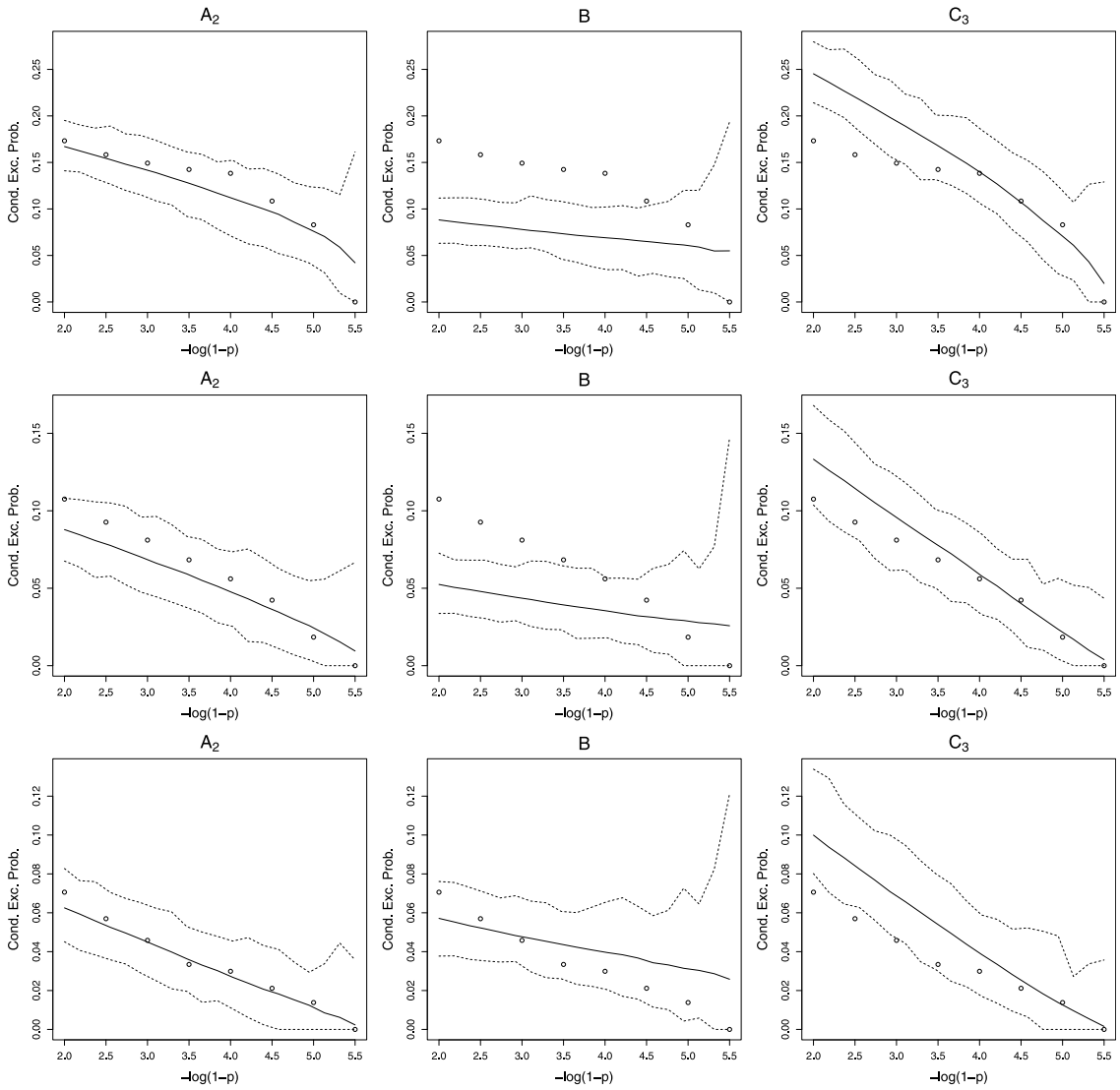


Fig. 10. Winter daily data: empirical and fitted values for the conditional probabilities $\Pr(Z(s) > z, s \in \mathcal{S} \mid Z(s_1) > z)$. The three columns correspond to models A_2, B and C_3, respectively. Top row: $\mathcal{S} = \{s_2, s_3, s_6, s_8, s_{10}\}$ (near sites data set); middle row $\mathcal{S} = \{s_{11}, s_{13}, s_{14}, s_{15}, s_{18}\}$ (medium sites data set); bottom row: $\mathcal{S} = \{s_{25}, s_{26}, s_{27}, s_{28}, s_{29}\}$ (far sites data set). The $1 - p$ values are such that $\Pr(Z(s_1) > z) = 1 - p$.

dependence is driven by a truncated extremal Gaussian max-stable process (Schlather, 2002) whereas at larger distances asymptotic independence is induced by any stochastic process with bivariate distributions satisfying a general condition proposed by Ledford and Tawn (1996). In this respect the hybrid models in Wadsworth and Tawn (2012) are particular instances.

Due to the intractability of the multivariate likelihoods parametric inference is carried out using a composite likelihood approach. A small and preliminary simulation study has shown that the inference procedure performs well, even when we have considered the boundary values for the mixture parameter.

In our real example we have highlighted that the max-mixture approach appears of interest for modeling environmental data. In particular it has the merit to overcome the limits of the max-stable models in which only asymptotic dependence or exact independence can be modeled.

Our attention has been concentrated on modeling the spatial dependence. In the future, we plan to consider spatio-temporal extensions that have fundamental interest in practice. Currently space–time models are still taking up little space in the literature and the major emphasis is in modeling asymptotic dependence treating the time just as additional

Table 3
Summary of the fitted models based on the daily exceedances from the Australian data. Standard errors are reported in parentheses.

Model	$\widehat{\rho}_1$	$\widehat{r}_1$	$\widehat{\rho}_2$	$\widehat{r}_2$	$\widehat{\beta}$	CLIC
A_1	78.71 (9.80)	833.76 (77.70)	1448.52 (57.72)	–	0.38 (0.02)	575 518.3
A_2	101.03 (13.93)	658.94 (54.26)	841.08 (51.23)	–	0.38 (0.02)	575 515.9
A_3	210.07 (10^{-13})	211.15 (10^{-13})	2164.57 (140.85)	1400.11 (95.08)	0 (10^{-13})	575 183.7
B	147.09 (6.17)	1706.55 (213.31)	–	–	–	580 455.
C_1	–	–	814.81 (19.34)	–	–	580 351.3
C_2	–	–	429.68 12.38	–	–	578 445.3
C_3	–	–	2084.84 (139.76)	1447.33 (106.76)	–	575 188.3

dimension of the space (Davis et al., 2013; Huser and Davison, 2014). However it seems reasonable to suppose that the spatial and temporal components behave asymptotically in a different way.

Acknowledgments

The research was partially supported by ANR-McSim and GICC-Miracle projects and by the Labex NUMEV. 137478, 2004. We are also indebted with Simone Padoan and the referees for comments that have led to improvements in the article.

References

Ancona-Navarrete, M., Tawn, J., 2002. Diagnostics for pairwise extremal dependence in spatial processes. Extremes 5, 271–285.
Bacro, J.N., Gaetan, C., 2012. A review on spatial extreme modelling. In: Porcu, E., Montero, J.M., Schlather, M. (Eds.), Advances and Challenges in Space–Time Modelling of Natural Events. Springer, New York, pp. 103–124.
Bacro, J.N., Gaetan, C., 2014. Estimation of spatial max-stable models using threshold exceedances. Stat. Comput. 24, 651–662.
Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. Statistics of Extremes: Theory and Applications. John Wiley & Sons, New York.
Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. Regular Variation. In: Encyclopedia of Mathematics and its Applications, vol. 27. Cambridge University Press, Cambridge.
Carlstein, A., 1986. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. Ann. Statist. 14, 1171–1179.
Casson, E., Coles, S.G., 1999. Spatial regression models for extremes. Extremes 1, 449–468.
Coles, S., Heffernan, J., Tawn, J., 1999. Dependence measures for extremes value analyses. Extremes 2, 339–365.
Cooley, D., Nychka, D., Naveau, P., 2007. Bayesian spatial modeling of extreme precipitation return levels. J. Amer. Statist. Assoc. 102, 824–840.
Davis, R.A., Klüppelberg, C., Steinkohl, C., 2013. Max-stable processes for modeling extremes observed in space and time. J. Korean Stat. Soc. 42, 399–414.
Davison, A.C., Gholamrezaee, M.M., 2012. Geostatistics of extremes. Proc. R. Soc. Lond. Ser. A 468, 581–608.
Davison, A.C., Huser, R., Thibaud, A., 2013. Geostatistics of dependent and asymptotically independent extremes. Math. Geosci. 45, 511–529.
Davison, A.C., Padoan, S.A., Ribatet, M., 2012. Statistical modelling of spatial extremes. Statist. Sci. 27, 161–186.
de Haan, L., 1984. A spectral representation for max-stable processes. Ann. Probab. 12, 1194–1204.
de Oliveira, T., 1962. Structure theory of bivariate extremes: extensions. Estud. Math. Estat. Econometrica 7, 165–195.
Engelke, S., Malinowski, A., Kabluchko, Z., Schlather, M., 2015. Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. J. R. Stat. Soc. Ser. B Stat. Methodol 77, 239–265.
Fawcett, L., Walshaw, D., 2007. Improved estimation for temporally clustered extremes. Environmetrics 18, 173–188.
Ferreira, A., de Haan, L., 2014. The generalized Pareto process; with a view towards application and simulation. Bernoulli 20, 1717–1737.
Gaetan, C., Grigoletto, M., 2007. A hierarchical model for the analysis of spatial rainfall extremes. J. Agric. Biol. Environ. Stat. 12, 434–449.
Guyon, X., 1995. Random Fields on a Network. Springer, New York.
Huser, R., Davison, A.C., 2014. Space–time modeling of extreme events. J. R. Stat. Soc. Ser. B 76, 439–461.
Jeon, S., Smith, R., 2012. Dependence structure of spatial extremes using threshold approach, Technical report. arXiv:1209.6344.
Kabluchko, Z., Schlather, M., de Haan, L., 2009. Stationary max-stable fields associated to negative definite functions. Ann. Probab. 37, 2042–2065.
Lavery, B., Kariko, A., Nicholls, N., 1992. A historical rainfall data set for Australia. Aust. Meteorol. Mag. 40, 33–39.
Ledford, A.W., Tawn, J.A., 1996. Statistics for near independence in multivariate extreme values. Biometrika 83, 169–187.
Ledford, A., Tawn, J., 1997. Modelling dependence within joint tail regions. J. R. Stat. Soc. Ser. B 59, 475–499.
Ledford, A., Tawn, J., 1998. Concomitant tail behaviour for extremes. Adv. Appl. Probab. 30, 197–215.
Lindsay, B., 1988. Composite likelihood methods. Contemp. Math. 80, 221–239.
Mardia, K.V., Watkins, A.J., 1989. On multimodality of the likelihood in the spatial linear model. Biometrika 76, 289–295.
Opitz, T., 2013. Extremal t processes: elliptical domain of attraction and a spectral representation. J. Multivariate Anal. 122, 409–413.
Padoan, S.A., 2013. Extreme dependence model based on event magnitude. J. Multivariate Anal. 122, 1–19.
Padoan, S.A., Ribatet, M., Sisson, S., 2010. Likelihood-based inference for max-stable processes. J. Amer. Statist. Assoc. 105, 263–277.
Resnick, S., 1987. Extreme Values, Regular Variation and Point Processes. Springer, New York.
Ribatet, M., Cooley, D., Davison, A., 2012. Bayesian inference for composite likelihood models and an application to spatial extremes. Statist. Sinica 22, 813–845.
Sang, H., Gelfand, A., 2010. Continuous spatial process models for spatial extreme values. J. Agric. Biol. Environ. Stat. 15, 49–65.
Schlather, M., 2002. Models for stationary max-stable random fields. Extremes 5, 33–44.

Schlather, M., Tawn, J.A., 2003. A dependence measure for multivariate and spatial extreme values: properties and inference. Biometrika 90, 139–156.
Serinaldi, F., Bárdossy, A., Kilsby, C.G., 2014. Upper tail dependence in rainfall extremes: would we know it if we saw it? Stoch. Environ. Res. Risk Assess. 29, 1211–1233.
Sibuya, M., 1960. Bivariate extreme statistics. Ann. Inst. Statist. Math. 11, 195–210.
Smith, R.L., 1990. Max-stable processes and spatial extremes, Preprint. University of Surrey.
Thibaud, E., Mutzner, R., Davison, A.C., 2013. Threshold modeling of extreme spatial rainfall. Water Resour. Res. 49, 4633–4644.
Varin, C., 2008. On composite marginal likelihoods. Adv. Stat. Anal. 92, 1–28.
Varin, C., Vidoni, P., 2005. A note on composite likelihood inference and model selection. Biometrika 92, 519–528.
Wadsworth, J., Tawn, J., 2012. Dependence modelling for spatial extremes. Biometrika 99, 253–272.
Wadsworth, J., Tawn, J., 2014. Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. Biometrika 101, 1–15.

Annexe B

G6 - A semiparametric method to simulate bivariate space-time extremes. AOAS (2017).

The Annals of Applied Statistics
2017, Vol. 11, No. 3, 1403–1428
DOI: 10.1214/17-AOAS1031

A SEMIPARAMETRIC METHOD TO SIMULATE BIVARIATE SPACE–TIME EXTREMES[1]

BY ROMAIN CHAILAN[*,†], GWLADYS TOULEMONDE[*] AND JEAN-NOEL BACRO[*]

University of Montpellier[*] and IBM France[†]

Coastal hazards raise many concerns, as their assessment involves extremely high economic and ecological stakes. In particular, studies on rarely observed but damaging events are quite numerous. In order to anticipate upcoming events of this kind, specialists need to extrapolate the results of their studies to events that have not yet occurred. Such events might be more extreme than those already observed and could therefore severely impact the coast. It is therefore paramount to propose methodologies to simulate such extreme conditions. Parametric and nonparametric statistical methods have already been used to assess environmental extreme quantities, from univariate framework to spatial context; however, they do not generally focus on the simulation of extreme environmental scenarios. This study introduces a semiparametric approach based on the Extreme Value Theory (EVT), dedicated to the simulation of extreme space–time processes. In the proposed application context, these processes describe near-shore hydraulic conditions. They nourish coastal impact models to assess hazards along the coast. The benefit of this approach is to be able to characterise coastal hazards on an event scale, meaning we can characterise the impact both in space and through time for a given extreme event. The usefulness of this space–time extreme modelling is illustrated by a risk analysis related to the long-shore impact of extreme wave events in the Gulf of Lions.

1. Introduction. Coastal hazards raise many concerns, as highly valuable economic and ecological assets are exposed along the world's coasts. Several studies demonstrate the significant benefits of understanding both littoral hydrodynamic and morphodynamic patterns in order to preserve them [e.g., Brunel et al. (2014), Gutierrez et al. (2015), Michaud et al. (2013)]. Some experts focus on extreme and devastating conditions, such as Campmas et al. (2014), who observes sediment transport patterns during the season of typhoons in Taiwan. Such a study helps preserve the littoral by enabling efficient beach nourishment.

An alternative to direct observations is the chaining of numerical models, which represent the physics from offshore to coastal areas. Typically, output data from

Received January 2016; revised February 2017.

[1]This work was supported by the french national program LEFE/INSU and by the labEx NUMEV.

Key words and phrases. Space-time extreme processes simulation, extreme value modelling, extreme waves, coastal hazards.

atmospheric and ocean circulation models force a wave model, which in turn feeds a littoral model [Bouchette et al. (2012), Michaud (2011)]. Refined output data of the latter are used to assess the hazard question.

In the case of observable extreme events, the reliability of physical models still holds. As soon as we consider very extreme events, their numerical simulation from physical models is generally unachievable. This is due to a lack of knowledge of boundary conditions and also of their physical reliability for such extreme quantities. As an alternative we propose to use statistical approaches, the main challenge being to extrapolate information from observations to simulate (very) extreme quantities.

From univariate to spatial approaches, analyses dealing with the understanding of extremes generally rely on the widely accepted Extreme Value Theory (EVT) [Beirlant et al. (2004), Coles (2001), Davison, Padoan and Ribatet (2012), Davison and Huser (2015)].

Various approaches have been presented to construct extreme scenarios of nearshore conditions like in Gouldby et al. (2014), but are generally not spatial. In the spatial context, Chailan et al. (2014) present an application of max-stable processes to analyse the spatial behaviour of extreme waves. The outputs of this study would be typical requirements to force physical hazard models in a coastal area. Indeed, max-stable processes are appealing in a spatial extreme context because they are the only possible nondegenerate limits for rescaled pointwise maxima of random processes [de Haan (1984)]. Inference of such max-stable processes is widely based on likelihood techniques, either in a frequentist approach [Engelke et al. (2015), Huser and Davison (2013), Padoan, Ribatet and Sisson (2010), Wadsworth and Tawn (2014)] or in a Bayesian one [Ribatet, Cooley and Davison (2012), Shaby (2014)]. Shaby and Reich (2012) present a Bayesian spatial extreme value analysis but interpreting the parameters in their hierarchical modelling is unfortunately not easy [for a possible interpretation as well as recent investigations on inference for spatial extremes, see Castruccio, Huser and Genton (2016)]. From a practical point of view, simulations of max-stable processes are of primary interest. They can be divided in two categories: unconditional and conditional simulations. For instance, Dieker and Mikosch (2015) propose exact simulations of the Brown–Resnick max-stable process at a finite number of locations. Their approach has been generalised by Dombry, Engelke and Oesting (2016) who also propose a more efficient algorithm. Wang and Stoev (2011) introduce a solution to construct a conditional process for max-linear processes. This work was extended by Bechler, Bel and Vrac (2015), Dombry and Eyi-Minko (2013), Dombry, Éyi-Minko and Ribatet (2013) in a less restrictive case. Nevertheless, the number of conditioning points remains limited and Lantuéjoul and Bel (2014) have recently remedied this weakness by introducing a new algorithm. However, since max-stable processes appear as natural for modelling block maxima (e.g., annual maxima), using simulations of such a process is more relevant in long-term questioning than in event-scale questioning due to the limited physical interpretation

of the simulated processes. This is clearly a limiting factor when questioning is more event-scale related (e.g., a submersion phenomenon along a coastline is an event-scale phenomenon and must be distinguished from a long-term problematic like the study of the decennial coastline dynamic). In the former case, not only the spatial information of an extreme process is needed, but also information characterising the time evolution of the analysed extreme event itself. For instance, this is essential in coastal engineering applications to compute dimensioning characteristics, such as the fatigue of seawalls through time when they are impacted by storm-waves.

In the following, we focus on space–time processes. Max-stable processes have also been developed and exemplified in a space–time context [Davis, Klüppelberg and Steinkohl (2013a, 2013b), Huser and Davison (2014), Embrechts, Koch and Robert (2016)] but are rarely alluded to in the literature. Their capacity to model complex dependence structures can still be questioned and the physical interpretation in any event-scale context of the simulated space–time processes issued by these models can be questioned as well.

Since these fully parametric methods do not directly answer the presented event-scale problematic and since it is unfeasible to model statistically the physical characteristics of storm events, we propose a methodology based on an empirical uplifting of real storms. This has the benefit of preserving the underlying physics of the considered processes. The idea is to exploit a peaks-over-threshold based approach and to propose a simulation scheme for extreme realisations. This does not assume any parametric model for the dependence structure. In the proposed methodology, we are focused on a semiparametric approach stemming from parts of the original work of Caires, de Haan and Smith (2011), de Haan and de Ronde (1998), Ferreira and de Haan (2014), Groeneweg, Caires and Roscoe (2012), summed up as follows.

Let $\{Z(s,t), s \in S, t \in \mathcal{T}_0\}$ be a space–time process, with $S \subset \mathbb{R}^2$ the area of interest and $\mathcal{T}_0 \subset \mathbb{R}^+$ the time dimension. In the sequel, such a process will represent an extreme event and will be named 'storm' for the sake of simplicity. The first step consists in selecting such a storm. To do so, the complete process is standardised in a preprocessing step. A combination between a preprocessing step and an extreme modelling has been proposed by Eastoe and Tawn (2009) but in a context of nonstationarity due to the presence of covariates. Here, a more extreme process is obtained by uplifting with a coefficient denoted $\zeta > 1$ the space–time process, which is initially transformed on a standard scale as $T(Z)$ where T is a marginal transformation detailed in Section 3.1. The process $T^{\leftarrow}(\zeta T(Z))$ becomes more extreme at the original scale.

Assuming that Z belongs to a max-stable domain of attraction, this approach is mathematically justified (see the Appendix). In practice, the space–time dependence structure of Z will be taken as constant in the extreme, leading to an asymptotic dependence context. Caires, de Haan and Smith (2011), Groeneweg, Caires and Roscoe (2012) use this methodology to simulate space–time extreme

processes. We leverage this approach to perform a bivariate simulation of such processes, with a view to better represent sea-states conditions at extreme levels. This leads us to develop a distinct strategy for the selection of storms and to use marginal distributions for the standardisation of the data as those used in Thibaud and Opitz (2015).

The behaviour of the produced storms is discussed around a case-study: the quantification of the long-shore mass flux of energy in a coastal area during extreme storms.

Since the presented methodology is applied to a large multidimensional volume of data, specific distributed algorithms are developed to process the data, which raises an additional technical dimension.

Section 2 introduces both the dataset used for this application and a preliminary study about the storms contained in it. Section 3 then presents in detail the statistical methodology and its justifications. The results are presented in Section 4.1 and then used for a risk analysis in Section 4.2. The final section provides a discussion about the introduced notions and their applications.

2. Data. Our region of interest is a semi-closed French coast area located in the northwestern Mediterranean sea, namely the Gulf of Lions (GOL) as presented in Figure 1. This study aims to simulate extreme space–time wave processes in order to use them as inputs for a littoral hazard model. For instance, a model of coastal submersion due to storm-waves, which is a physical process depending on near-shore hydrodynamic conditions. Such a model is forced by inputs describing

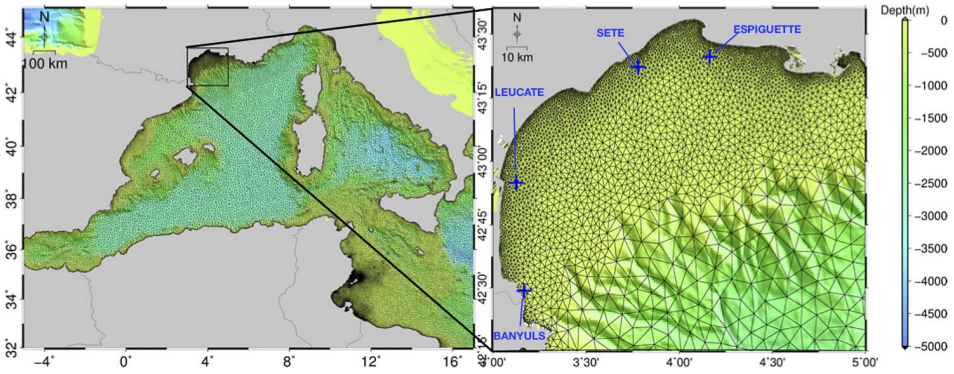


FIG. 1. *The left panel is the full extension of the domain considered for the hindcast. The right panel is the studied area*: *the Gulf of Lions* (*GOL*). *The crosses indicate the locations of surface buoys measuring waves features. The colour scale indicates the bathymetry*, *that is*, *underwater topography*, *of the northwestern Mediterranean sea. The computational mesh used for the hindcast is also overlaid. It is composed of* 47,086 *nodes with a spatial resolution ranging from* 1 *km to* 12 *km. The right panel is a zoom of the grid on the GOL. Computational nodes situated in the GOL form the set* $\mathcal{M}$.

the sea-states conditions at an instant t. Generally, these inputs are the mean wave direction $\psi(t)$, the significant wave height $\mathrm{Hs}(t)$ and the peak wave period $\mathrm{Tp}(t)$.

Three sources are principally considered in obtaining such data. The first is surface-buoys that monitor these three variables. In the GOL, there are four surface-buoys as illustrated in Figure 1. These observations are accurate but sparsely provided in our region of interest, in both space and time dimensions. Spatial scarcity would degrade the spatial modelling of the process whereas short time series would degrade the quality of the extrapolation to more extreme values.

An alternative is to use satellite-altimeter datasets. The major issue is that only $\mathrm{Hs}(t)$ can be observed from an altimeter. Satellites embedding Synthetic Aperture Radar (SAR) must be considered if wave direction and wave period are required, but the time series are shorter (first launch in the 1990s). Moreover, since satellites tracks are nonregular through time and space around the globe, any extreme statistical analysis considering such datasets [see, e.g., Raillard, Ailliot and Yao (2014)] becomes hard to handle, especially when the modelling concerns events in a fixed and relatively confined area.

The final way to observe wave data variables is the use of the numerical simulation. Chailan et al. (2014) proposed a 52-year hindcast of wave features over the north-western Mediterranean sea, extending from the Strait of Gibraltar to the south of Italy. This hindcast is obtained by the use of a widely recognised wave numerical model in ocean community. In the sequel, this hindcast—validated against in situ observations—is used since it provides the longest and refined wave time series for this area to the best of our knowledge [Chailan (2015), Chapter 3]. Details are given in the next subsection.

2.1. *A 52-year wave hindcast.* The 52-year hindcast is produced with the WAVEWATCHIII® v4.18 (WW3) wave model [Tolman (2014)]. Two regional reanalyses have been used as forcing conditions—meaning used as inputs of the numerical model: Herrmann and Somot (2008) for atmospheric conditions and Herrmann et al. (2010) for ocean conditions. The bathymetry used has a spatial resolution of 0.0083 degree. The physical time range of the simulations ranges from January 1961 to December 2012 at an hourly scale. Finally, the unstructured computational grid illustrated in Figure 1 is composed of 47,086 nodes—3,944 for the GOL only—with a spatial resolution ranging from 1 km to 12 km.

The former quantities of interests [i.e., $\psi(t)$, $\mathrm{Hs}(t)$ and $\mathrm{Tp}(t)$] derive from the computed wave spectral density at each node of the mesh. For the GOL only, these three variables are stored in a binary file of 19 GB.

The dataset produced is validated against the records of the four surface buoys, at a yearly scale in terms of the time series available. As it is often the case, the wave model shows a good performance but tends to slightly underestimate the extreme occurrences. One way to understand the performance of the hindcast is to look at both marginal and joint measures of validation.

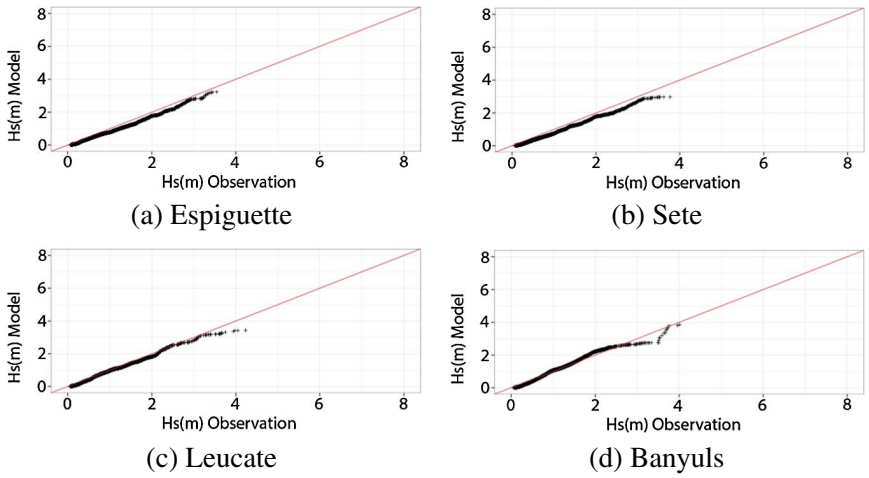


FIG. 2. *Quantile–quantile plots of the observed significant wave heights* (Hs) *against the modelled ones for* 2012. *Locations are the four littoral surface buoys of the GOL.*

For instance, the median over all buoys of the yearly correlation factors reaches 0.903 while the median of the root mean square errors is 0.272. Figure 2 illustrates quantiles of the observed significant wave heights (Hs) against the modelled ones for the year 2012. For this year, the former measures approximate their medians for each location, respectively. It makes 2012 a representative candidate to diagnose the overall hindcast quality [see Chailan (2015), Chapter 3, for additional measures].

The observed bias might not come from the wave model only [Rascle and Ardhuin (2013)]. Indeed the forcing re-analyses, especially the wind fields, are sometimes underestimated for instantaneous and abrupt wind gusts. Consequently, the generated wind-waves are underestimated as well. Despite these slight underestimations, the produced data are relatively satisfactory.

Insofar as it is a key feature of our study, the performance of the numerical model in regards to the spatial dependence structure must be presented as well. An analysis of joint survival probability is performed to this end. The purpose is to compare from each source—either buoys or numerical models—the joint probability of exceedance from various sets of sites corresponding to the locations of the buoys. In the empirical computations and for each sub-set of locations, the records taken into account are those that are simultaneously available at each site of the set. The thresholds quantiles are calculated, respectively, for each source of data. This limits misinterpretation due to bias from marginal intensities.

For the sake of clarity, only three out of the ten combinations available (four buoys) are presented in Figure 3, but similar results are observed regarding the other sets.

Those plots reveal a good match between survival joint probabilities from buoys compared to the ones from the numerical model, whatever the distance between the

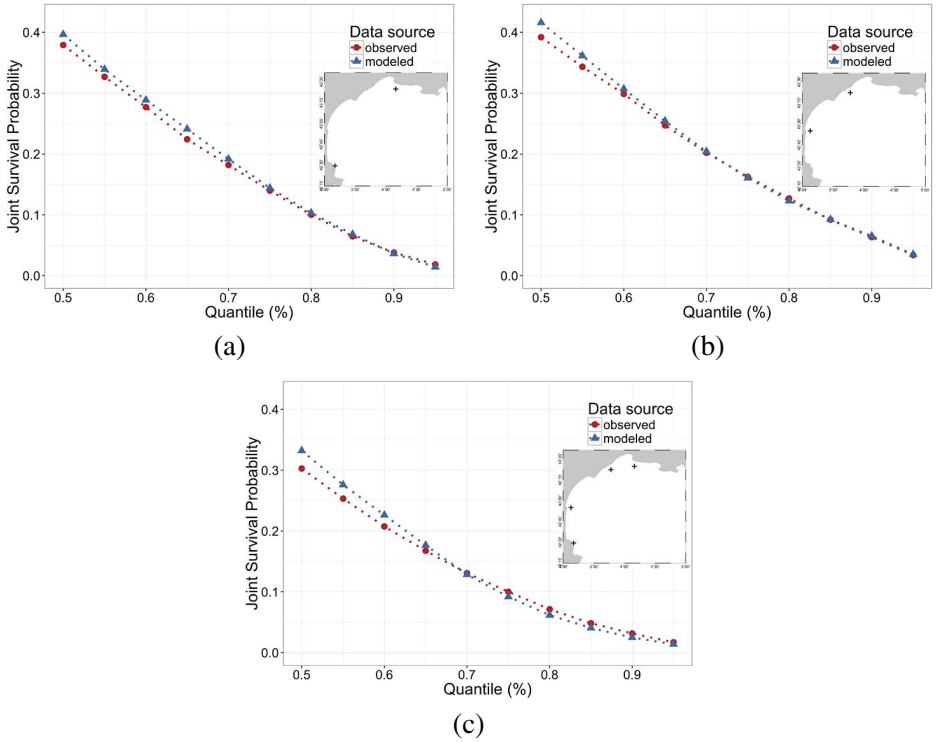


FIG. 3. *Joint survival probabilities of exceedance of significant wave heights* (Hs). *The empirical probabilities are computed from each data source* (*buoys or numerical model*). *Each sub-panel represents joint probabilities over various sets of sites corresponding to the buoys' locations. Selected sites are localised by the crosses on the map for each sub-panel.*

sites or their numbers. The adequacy is especially valid for joint probabilities of exceedance over high quantiles but with higher bias on smaller quantiles. It means that small waves are more spatially structured when observed from the numerical wave model but the spatial dependence structure is properly modelled for high waves. This remark reinforces the relevance of considering those produced data as observations in the sequel.

2.2. *Preliminary analysis.* A preliminary analysis is realised to develop our expertise on the wave data previously presented. As the reader may know, wind is the major factor of wave construction. The GOL is exposed to three dominant wind regimes. The first two are called *Tramontane* and *Mistral*. They come from the northwest and north, respectively. The last is called *Marin* and comes from the southeast. When the region is exposed to a Tramontane or Mistral episode or both, waves tend to propagate towards the southeast but are formed far from the coastline. This is due to a too short *fetch* zone—the zone where the wind stresses the sea-surface causing the growth of the waves. On the contrary, as soon as the

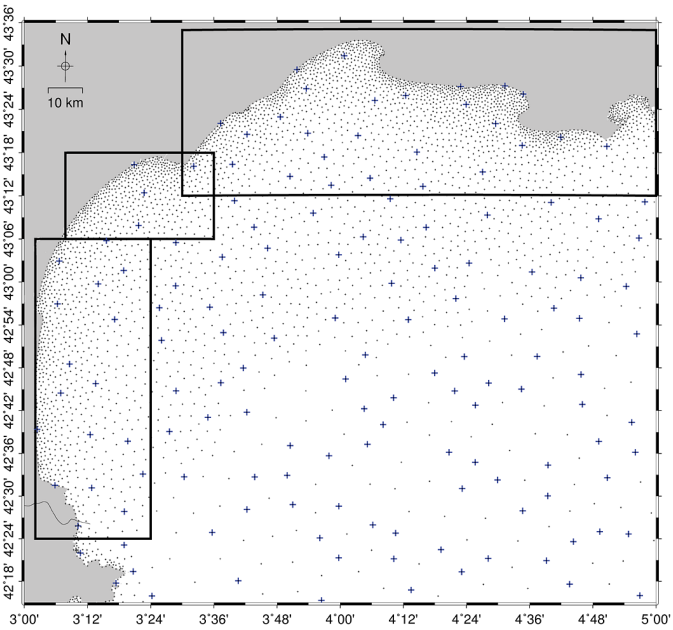


FIG. 4. *Spatial specification. Littoral area $S^* \subset S$ is the union of squared areas. From expert advice, if* Hs *is high in S^* the coastline is likely to be impacted. Wave data are available at the set of locations of the mesh nodes in this area, which is denoted $\mathcal{M}^* \subseteq S^*$. Cross points form a subset χ of* 140 *sites from the locations of the computational mesh nodes. χ is constructed in manner of spatially representing all observation locations.*

area is exposed to a Marin episode, waves are formed offshore and are propagated to the coasts. In such cases, the waves impact the coastline. Winds hitting the GOL are sometimes more complex and the resulting hydrodynamic is fairly modified: occasionally a southwest wave-flux is dominant in the GOL. Experts advise that the relevant storms to study the impact on the coastline are those in which the Hs variable reaches high values inside a very littoral area denoted S^*. For the GOL, we decided to choose the union of the determined areas (Figure 4).

Beside these physical characteristics, some statistical information can provide valuable information about the general behaviour of a wave-storm in the GOL. In particular, the extremal coefficient θ [Smith (1990), Schlather and Tawn (2003)] is a quantity that enables us to quantify the dependence in the context of extreme values.

This measure stems from the following reasoning. Without loss of generality, let us consider identically distributed random variables $Y^{(1)}, \ldots, Y^{(M)}$ with unit Fréchet distribution, that is, $P(Y^{(i)} \leq y) = e^{-1/y}, i = 1, \ldots, M, 0 < y < \infty$. If the joint distribution of $(Y^{(1)}, \ldots, Y^{(M)})$ is a multivariate extreme value distribution, it is well known that the joint probability $P(Y^{(1)} \leq y, \ldots, Y^{(M)} \leq y)$ can be expressed as $e^{-\theta/y}$. The so-called extremal coefficient $\theta = \theta(Y^{(1)}, \ldots, Y^{(M)})$,

$1 \le \theta \le M$ summarises the extremal dependence. The limiting case $\theta = 1$ represents the full dependence whereas $\theta = M$ represents the total independence.

In the context of threshold-based extreme value methods, realisations above a high threshold are considered as extreme. Assuming predetermined thresholds vectors $(u_j^{(1)}, \ldots, u_j^{(M)})$ and random vectors $(Y_j^{(1)}, \ldots, Y_j^{(M)})$, $1 \le j \le N$, the $Y_j^{(k)}$ are observed only if $Y_j^{(k)} > u_j^{(k)}$; otherwise, $Y_j^{(k)}$ is censored at $u_j^{(k)}$.

In this context, Smith in Caires, de Haan and Smith (2011) defines a natural estimator of the extremal coefficient function θ as

$$\hat{\theta} = m \Big/ \sum_{j=1}^{N} \frac{1}{\max(Y_j, u_j)}, \tag{2.1}$$

where Y_j and u_j are defined as $\max(Y_j^{(1)}, \ldots, Y_j^{(M)})$ and $\max(u_j^{(1)}, \ldots, u_j^{(M)})$, respectively; m is the number of excesses $Y_j > u_j$.

The pairwise extremal coefficient is commonly considered in statistical applications, meaning $Y_j = \max(Y_j^{(1)}, Y_j^{(2)})$ with $M = 2$ in (2.1). In the sequel, three extremal coefficients are introduced and estimated for the sea-states hindcast dataset. The first two are related to the dependence of the variable Hs through time and spatial distance, respectively. The time extremal coefficient $\theta^{\mathrm{tim}}(k)$ measures the dependence between pairs of observations of Hs separated by a time lag k, at a given location. The spatial extremal coefficient $\theta^{\mathrm{spa}}(h)$ measures the dependence between pairs of Hs observations separated by a spatial distance h, at a given time.

Figure 5 presents the extremal coefficients estimated for the full period (1961–2012) of the hindcast on a yearly block of data in order to monitor their fluctuations. Here, u_j in (2.1) is set as a 0.95-quantile to avoid issues stemming from a lack of data. Figure 5(a) presents the estimations $\widehat{\theta^{\mathrm{spa}}}(h)$ for two locations separated by a distance h. In this case, $(Y_j^{(1)}, Y_j^{(2)})$ in (2.1) corresponds to $(Y(t_j, s), Y(t_j, s + h))$. To compute $\widehat{\theta^{\mathrm{spa}}}(h)$, only a subset χ of 140 sites (Figure 4) from the computational mesh is considered. It limits the combinations of pairs available in the dataset. The selection of sites is optimised to fairly cover the entire area as described in Chailan et al. (2014). Estimations $\widehat{\theta^{\mathrm{spa}}}(h)$ are binned to 1,500 distinct distances h.

We observe that $\widehat{\theta^{\mathrm{spa}}}(h)$ is always strictly inferior to 2. More precisely, it is approximately 1.75 at the longest distance, exemplifying that dependence within a storm on the GOL, which is a relatively confined area, seems to be relatively persistent even at the longest distances. Beside the presented omnidirectional graphic, directions of pairs were considered and regrouped to compute the directional estimation of the dependence structure. This did not demonstrate a clear anisotropic pattern and, therefore, graphics are not presented here.

Figure 5(b) presents the estimations of $\theta^{\mathrm{tim}}(k)$ for pairs separated by a time lag k. In this case $(Y_j^{(1)}, Y_j^{(2)})$ in (2.1) represents $(\max_{s \in \mathcal{M}^*}(Y(t_j, s)), \max_{s \in \mathcal{M}^*}(Y(t_j + k, s)))$, with $\mathcal{M}^*$ the observation locations situated in S^* the

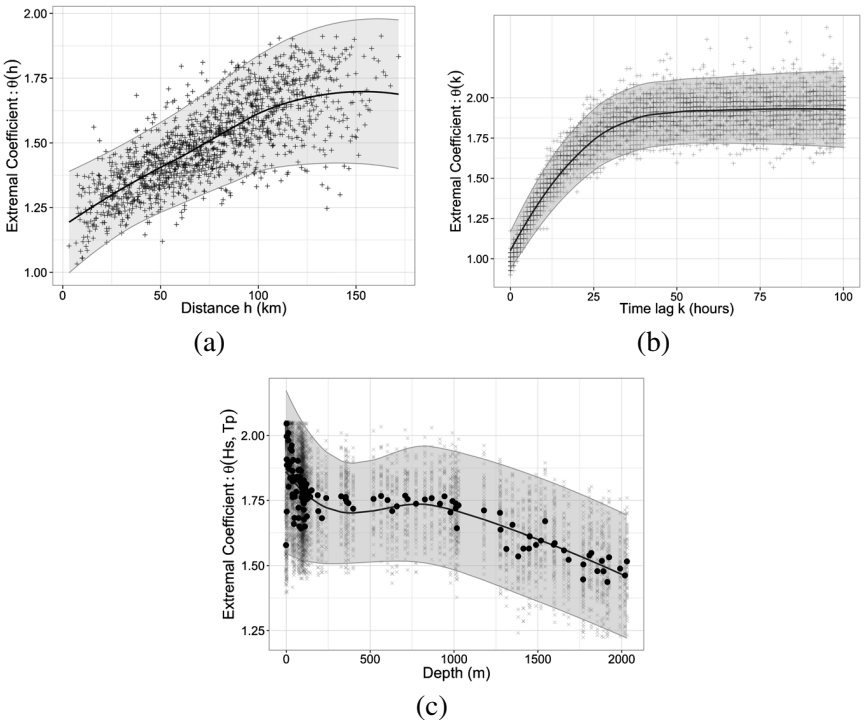


FIG. 5. *Estimations of the three extremal coefficients* (*see text for details*). *For each pair, the coefficients are estimated for the full period* (1961–2012) *of the hindcast on yearly block of data*. (a) *The extremal coefficients* $\theta^{\mathrm{spa}}(h)$ *estimated on* χ *from pairs of* Hs *values separated by a distance* h *given in kilometers*. *Estimations are binned to* 1,500 *distinct distances* h. (b) *Extremal coefficients* $\theta^{\mathrm{tim}}(k)$ *are estimated from pairs of* Hs *values separated by a lag* k *in hour.* (c) *Extremal coefficients* $\theta(\mathrm{Hs}, \mathrm{Tp})$ *estimated from the significant wave height* Hs *and the peak wave period* Tp *at locations* $s \in \chi$ *are ordered by their corresponding bathymetry. The dots are the median values from estimated pairwise coefficients*. *In each sub-panel, the straight line and its shadow envelope are respectively a fitted polynomial regression model and its* 95% *prediction interval*.

very littoral zone presented above. The arbitrary choice of S^* is still related to the final goal of the document: quantifying coastal hazards. With such littoral areas, only storms impacting the shoreline area are considered in the measure. We can observe from Figure 5(b) that $\widehat{\theta^{\mathrm{tim}}}(k)$ narrows 1.9 and becomes almost steady at $k = 50$. Hence, we can state that the dependence within a storm impacting the littoral will be considered as persistent only up to 50 hours.

The proposed uplifting procedure relies on a crucial hypothesis which is max-stable context. Indeed, we assume that the space–time dependence structures are constant in the extreme. Figures 5(a) and 5(b) show that this hypothesis is reasonable with our data, when considering a time lag smaller than 50 hours, corresponding to an extremal coefficient strictly inferior to 2.

Finally, to assess the dependencies between the two wave variables Hs and Tp observed at the same time and at the same location, we consider a third extremal coefficient $\theta(\mathrm{Hs}, \mathrm{Tp})$. Let $\mathrm{Hs}(t_j, s)$ and $\mathrm{Tp}(t_j, s)$ denote the significant wave height and the peak wave period at time t_j and location s, respectively. In this case, $(Y_j^{(1)}, Y_j^{(2)})$ in (2.1) represents $(\mathrm{Hs}(t_j, s), \mathrm{Tp}(t_j, s))$. Estimation $\widehat{\theta}(\mathrm{Hs}, \mathrm{Tp})$ is computed using the data from the subset χ. Figure 5(c) illustrates such estimation. By ordering the estimated bivariate extremal coefficients by the depth of the observation sites, we show that the deeper the sites, the more Hs and Tp remain dependent within their extreme realisations. In general, we can deduce that those two variables are fairly dependent, with an extremal coefficient inferior to 2, even if the waves mechanic may behave differently in very shallow waters.

3. Semiparametric storm uplifter.

3.1. *Extreme space–time processes.* In the sequel, $\{X(s,t), s \in S, t \in \mathcal{T}\}$ denotes a random space–time process with S a compact subset of $\mathbb{R}^d$ and $\mathcal{T}$ a compact subset of $\mathbb{R}^+$. Such a random process represents a random variables collection indexed by both space and time which is in the space of continuous real functions on $S \times \mathcal{T}$ denoted $C(S \times \mathcal{T})$. We suppose that $\{X(s,t), s \in S, t \in \mathcal{T}\}$ is in the domain of attraction of a max-stable process [de Haan and Lin (2001), de Haan and Ferreira (2006)]. In other words, we suppose that there exist continuous functions $a_n(s,t)$ positive and $b_n(s,t)$ such that the process

$$\left\{ \max_{1 \le i \le n} \frac{X_i(s,t) - b_n(s,t)}{a_n(s,t)} \right\}_{(s,t) \in S \times \mathcal{T}}$$

with $X_1, \ldots, X_n$ independent copies of X, converges in distribution to a max-stable process η in $C(S \times \mathcal{T})$. Since convergence of marginals and convergence of dependence structure can be split up, we consider, in the sequel, the standardised process $1/(1 - G_{X(s,t)}(X(s,t)))$ where $G_{X(s,t)}$ corresponds to the distribution of $X(s,t)$. Such a process has marginal standard Pareto distributions and belongs to the domain of attraction of the unit Fréchet distribution. Following Thibaud and Opitz (2015), it is convenient to fix a high threshold function $u(s,t)$ and to assume that the marginal distributions of this process satisfy

$$P\big(X(s,t) > x\big) = \big[1 + \xi(s,t)\big(x - \mu(s,t)\big)/\sigma(s,t)\big]_+^{-1/\xi(s,t)}, \tag{3.1}$$

for $x > u(s,t)$, with real parameters $\mu(s,t) < u(s,t)$, $\sigma(s,t) > 0$ and $\xi(s,t)$, such that the right-hand side of (3.1) is less than unity.

As a consequence, to result in a process with standard Pareto margins, we can define the standardised process X^* as follows:

$$X^*(s,t) = T\big(X(s,t)\big) = \big[1 + \xi(s,t)\big(X(s,t) - \mu(s,t)\big)/\sigma(s,t)\big]^{1/\xi(s,t)}. \tag{3.2}$$

3.2. *Method*. As presented in the Introduction, the outline of the methodology consists of four steps. First, data are marginally transformed. This enables us to manipulate the data on a standard scale. Here, we use a transformation to reach the standard Pareto scale. Then we need to extract storms from the dataset. Once storms are extracted, the data are uplifted to higher values, with a control on the marginal amplification coefficient. Finally, the data are transformed back to their original scale by inverting the transformation. Details of these four steps of the presented methodology are given in this subsection.

The first step consists in standardising $X(s,t)$ to a standard Pareto scale according to (3.2). In practice, the parameters are unknown and need to be estimated. In this first approach, we suppose the threshold and the parameters to be constant over time, depending only on space. One can alternatively use more sophisticated expressions of those quantities to deal with a potential nonstationarity of the process, for example, seasonality and directional effects might be better explained doing so [e.g., Jonathan, Ewans and Randell (2013)].

In each site, parameter estimations $\hat{\mu}(s)$, $\hat{\sigma}(s)$, $\hat{\xi}(s)$ are obtained by the maximum likelihood method using data above a high threshold $u(s)$ which can be chosen as a high quantile for a fixed s (here the 0.99-quantile). Since marginal data may have some short-term dependences, they are de-clustered before being used to estimate the parameters [Coles (2001), Section 5.3.2]. In this paper, the de-cluster procedure has been configured with an interval of 5 consecutive values below u_s to consider an exceedance as a new cluster, that is, 6 hours after the last exceedance. This step allows us to reach the independence condition assumed in the estimation procedure. Using such estimators in (3.2), let $\{\tilde{X}^*(s,t), s \in S, t \in \mathcal{T}\}$ denote the obtained standardised process. Note that this preprocessing step relies on different techniques from those used in Caires, de Haan and Smith (2011), Groeneweg, Caires and Roscoe (2012).

The second step consists in extracting storms on a standardised scale from the data. To extract the biggest storm, the maximum value of $\widetilde{X}^*(s,t)$ is searched over the subset of sites $\mathcal{M}^*$, which might be a single reference location, locations of the entire space S or locations of some area in between. This point leads to a distinct strategy of selection of storms from Caires, de Haan and Smith (2011), Groeneweg, Caires and Roscoe (2012). Let us assume this maximum occurs at time t_1. We fix the total storm duration as 2δ. Consequently, such a storm is a subset in the time dimension of $\{\tilde{X}^*(s,t), s \in S, t \in \mathcal{T}\}$, therefore, defined as $\tilde{Z}^* = \{\tilde{X}^*(s,t), s \in S, t \in \mathcal{T}_0 \subset \mathcal{T}\}$. For this first storm, $\mathcal{T}_0 = [t_1 - \delta, t_1 + \delta]$.

The period $\mathcal{T}_0$ is hidden from the selection of the second biggest storm. Furthermore, we introduce a time value which is a "precaution time-lag" ε to insure the independence of the storms. The selection of the second biggest storm will consist in identifying the maximum value of $\widetilde{X}^*(s,t)$ over the subset of sites $\mathcal{M}^*$ with $t \in \mathcal{T} \backslash [t_1 - \delta - \varepsilon, t_1 + \delta + \varepsilon]$. The two values δ and ε are generally defined according to expert advice or from preliminary analyses or both. In this study, the specific values of these parameters are given and explained in Section 4.1.

Algorithm 1: Storm selection

Input : $\{\tilde{X}^*(s,t),\ s \in S,\ t \in \mathcal{T}\}$, space–time observations on a standard scale.
p' the maximum number of storms to select.

Output: $\{\tilde{Z}_i^*, i \in \{1,\ldots,p\}\}$ with $p \le p'$, a sorted collection of i.i.d. storms

begin
  $i = 1$, $\delta \leftarrow$ Cst, $\varepsilon \leftarrow$ Cst, $T \leftarrow \mathcal{T}$, $T' \leftarrow T$;
  while $(i \le p')$ and $(\max_{s\in\mathcal{M}^*, t\in T'} \tilde{X}^*(s,t) > 1)$ **do**
    $t_i \leftarrow \arg\max_t\{\tilde{X}^*(s,t)\}$; // $s \in \mathcal{M}^* \subseteq S$ and $t \in T'$.
    $\tilde{Z}_i^* \leftarrow \tilde{X}^*(\cdot, t)$ with $t \in T \cap [t_i - \delta,\ t_i + \delta]$;
    $T' \leftarrow T' \setminus [t_i - \delta - \varepsilon,\ t_i + \delta + \varepsilon]$;
    $i = i + 1$;
  return $\{\tilde{Z}_1^*, \tilde{Z}_2^*, \ldots, \tilde{Z}_p^*\}$;

The general iterative scheme to select storms is presented in Algorithm 1. It is noticeable that the stop condition of the algorithm implies that there is at least one exceedance of the site marginal threshold in each selected storm. The algorithm would select storms until the required and arbitrary number of storms p' is reached or until the exceedance condition is no longer satisfied.

Finally, let $\{\tilde{Z}_i^*(s,t), i \in \{1,\ldots,p\}\}$ denote a collection of such space–time processes and represent the p highest storms available in the transformed dataset.

It is relevant to compare them with each other in term of their extremeness. In the sequel, the definition of extremeness of a so-called storm $\{Z^*(s,t), s \in S, t \in \mathcal{T}_0 \subset \mathcal{T}\}$ relies on the level corresponding to the within-storm maxima $z_{\max} = \max_{s,t}\{Z^*(s,t), s \in \mathcal{M}^* \subset S, t \in \mathcal{T}_0 \subset \mathcal{T}\}$. Consequently, a storm $\{Z_1^*\}$ is considered more extreme than $\{Z_2^*\}$ if $z_{1,\max} > z_{2,\max}$.

In extreme value theory, a return period m is associated with a return level r_m. The return level r_m is reached once over the return period m in mean. By definition, this is no more than the $(1 - \frac{1}{m})$-quantile of the block maximum distribution. We define the return period of a storm $\{Z^*(s,t)\}$ as equal to the marginal return period associated with the within-storm maxima $z_{\max}$ observed at location $s_{\max}$. The location $s_{\max}$ is either fixed as a reference site or defined as equal to $\operatorname{argmax}_{s\in\mathcal{M}^*}\{\tilde{Z}^*(s,t)\}$.

The third step consists of an uplifting technique. To obtain more severe storms (with a longer return period), processes $\tilde{Z}_i^*, i \in \{1,\ldots,p\}$ are multiplied by a coefficient factor superior to unity and denoted ζ_i. The coefficient ζ_i is applied to the entire duration of the storm i. Hence, $\zeta_i \tilde{Z}_i^*(s,t), \zeta_i > 1, i \in \{1,\ldots,p\}$, is the collection of the uplifted storms at the standardised scale.

For the final step, each uplifted storm is transformed back to its original scale by

$$\tilde{Z}_i(s,t) = T^{\leftarrow}\big(\zeta_i \tilde{Z}_i^*(s,t)\big), \qquad i \in \{1,\dots,p\}, \tag{3.3}$$

where $T^{\leftarrow}(Y(s,t)) = \hat{\mu}(s) + \hat{\sigma}(s)\frac{[Y(s,t)]^{\hat{\xi}(s)}-1}{\hat{\xi}(s)}$.

We obtain here a collection of heavier extreme storms from a set of observed extreme storms.

It is important to highlight that an observed extreme storm $Z_i^*(s,t)$ is defined if and only if

$$\max_{s\in\mathcal{M}^*} Z_i^*(s,t) > 1, \tag{3.4}$$

meaning that there is at least one exceedance of the site marginal threshold. This uplifting proposition relies on a mathematical justification given in the Appendix. In this detailed proof, it has been shown that there is actually no limitation in uplifting bivariate processes $\{Z_{1,i}^*, Z_{2,i}^*\}$ conditioned to (3.4) is satisfied for one of the margin.

What is further remarkable is that such a uplift method of a space–time process appears as naturally linked to the GPD process framework. This framework was initially introduced by Ferreira and de Haan (2014). Dombry and Ribatet (2015) generalise this result by considering conditional events characterised through a continuous and homogeneous risk function $\ell(\cdot)$. The case from Ferreira and de Haan (2014) corresponds to $\ell(f) = \sup_{s\in S} f(s)$ and the ℓ function we are considering here corresponds to $\ell(f) = \max_j f(s_j,t)$. As a consequence, the limit of the conditional distribution we consider corresponds to the distribution of a GPD process.

Other remarks can be made with regard to the construction of the processes. First, note that in (3.3), the coefficient ζ_i relative to the uplifted storm i can be chosen in several ways as long as it is superior to 1.

We can consider, and this is in fact the choice we made, the special case $\zeta_i = \frac{T(z_m)}{T(z_{\max})}$, where $z_{\max}$ is still the within-storm maxima and z_m is the return level corresponding to the m-year return period at location $s_{\max}$. Implemented in Groeneweg, Caires and Roscoe (2012), Smith in Caires, de Haan and Smith (2011) interprets such a transformation as an uplift from a storm with a given return period to a storm with a return period equal to m. In that case, ζ_i is obviously storm-dependent and this choice enables us to uplift different storms to a comparable level. However, other choices for ζ_i could be proposed, for example, in Caires, de Haan and Smith (2011), de Haan proposes another approach which can be interpreted as an uplifting of the threshold of the peaks-over-threshold process Z_i. As another example, ζ_i, $i = 1,\dots,p$ could be obtained as independent realisations of a standard Pareto distribution. In that case, our approach should be very similar to the constructive representation of the Pareto process proposed by Dombry and Ribatet (2015). To the best of our knowledge, there are few results about simulations of GPD processes and consequently our results may also be of interest.

4. Results.

4.1. *Uplifted storms*. The presented method is applied to the 52-year sea-states condition dataset. To cope with the computational demand of dealing with nearly 4000 locations, algorithms are implemented in a dedicated R code and parallelised via the Message Passing Interface (MPI) protocol. All computations are performed on a cluster composed of 96 cores, which reduces the overall computation duration to nearly 5 hours.

From this point on and for the sake of simplicity, the definition of storm embraces the multivariate space–time processes composed of Hs, Tp and directions ψ.

We worked on the 10 highest storms observed to uplift both Hs and Tp variables, resting on the proposed bivariate approach. In our case study, Hs is the variable that conditions the bivariate space–time processes selection. It avoids selecting events with high Tp but low Hs, a phenomenon that can be observed in nature. Consequently, only highly energetic wave processes are considered because at least one component in $\mathcal{M}^*$ exceeds its threshold. In this application, marginal thresholds correspond to marginal 0.99-quantiles.

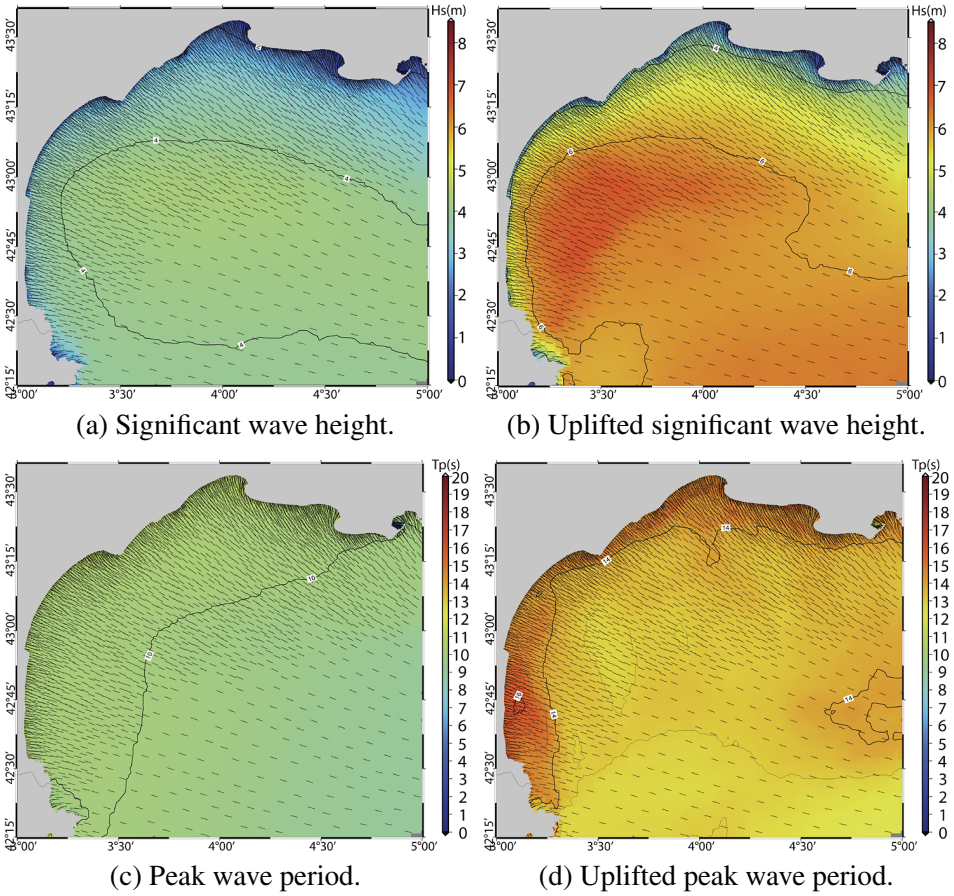
We are concerned with modelling storms that impact the coastline only. Hence, we chose to set S^* equal to the coastline-band area illustrated in Figure 4. This restriction in the storm detection area prevents the selection of offshore storms that do not propagate to the coast in the execution of Algorithm 1.

From the preliminary analysis in Section 2.2, we determine that storms last about 50 hours: the duration for which the extremal coefficient appears to be steady, revealing a persistence of the dependence structure within a storm up to that time. Thus, the selected value of δ is equal to 24 (hours). To select only i.i.d. storms, the value of ε is also equal to 24 (hours). This parameter is set to avoid the selection of overlapping storms. In this application, it would have been set to 0 without any consequence since no overlaps had been detected in this configuration.

Both $\zeta_{i,\mathrm{Hs}}$ and $\zeta_{i,Ts}$ are chosen to uplift original storms to m-year return period storms following the implementation of Groeneweg, Caires and Roscoe (2012). It is remarkable that any uplifted storms in the same return period might be compared to realisations of the distribution of the storms at this return period. Hence, having the control on the return period of storms is the easiest way to interpret and compare the impacts of storms from a coastal engineering point of view. In this application, $s_{\max}$—the within storm maxima—is chosen among the entire set $\mathcal{M}^*$ of locations available in the littoral area. The location $s_{\max}$ might be different for the two variables. Figure 6 illustrates one of the uplifted storms.

Note that mean wave directions are conserved during the uplift procedure.

Among the set of 10 scenarios, the variability of the fields observed are quite large, but are unsurprisingly dominated by fluxes from the south, southeast or east. This is a direct consequence of choosing $\mathcal{M}^*$ as a very littoral area.



(a) Significant wave height. (b) Uplifted significant wave height.

(c) Peak wave period. (d) Uplifted peak wave period.

FIG. 6. *Comparison of a storm uplifted to its* 100*-year return period, at its peak. The left panels illustrate the original storm; the right panels illustrate the uplifted storm. The arrows indicate the mean wave directions.*

4.2. *Uplifted storms at work: A risk analysis.* Coastal hazards such as submersion, erosion or beach contamination are usually quantified from formulae that require the computation of mass flux of energy towards the shoreline, given off the shoaling zone where waves do not interact significantly with the sea bottom. We usually distinguish between cross-shore and long-shore contributions, depending upon the goal of the application. For instance, the calculation of the alongshore-sand transport [Bagnold (1966), CERC (1984)] requires the long shore mass flux of energy. In the following, we strictly consider the long- shore impact ϕ of the deep water mass flux of energy $\mathcal{Q}$ to the shoreline, which is a relevant expression to tackle any analysis of shoreline dynamics. We model evolution of such a quantity during extreme wave storms.

For a given storm event S, we compute the impact $\phi_{i,t}^{(S)}$ at a location $c_i \in \mathcal{C}$ and at a time t of the mass flux of energy $\mathcal{Q}_{i,t}$ coming from waves at a location $l_i \in \mathcal{L}$ (see Figure 7). The long-shore impact is calculated by

$$\phi_{i,t}^{(S)} = \mathcal{Q}_{i,t} \sin(\omega_{i,t}) \cos(\omega_{i,t}), \tag{4.1}$$

where $\omega_{i,t}$ represents the angle of the wave propagation at l_i at a time t and is function of the wave direction $\psi_{i,t}$.

Practically, $\mathcal{Q}$ is derived from the variables Hs, Tp characterising the sea-state conditions at various points along an iso-bathymetric baseline. Such a mass flux of

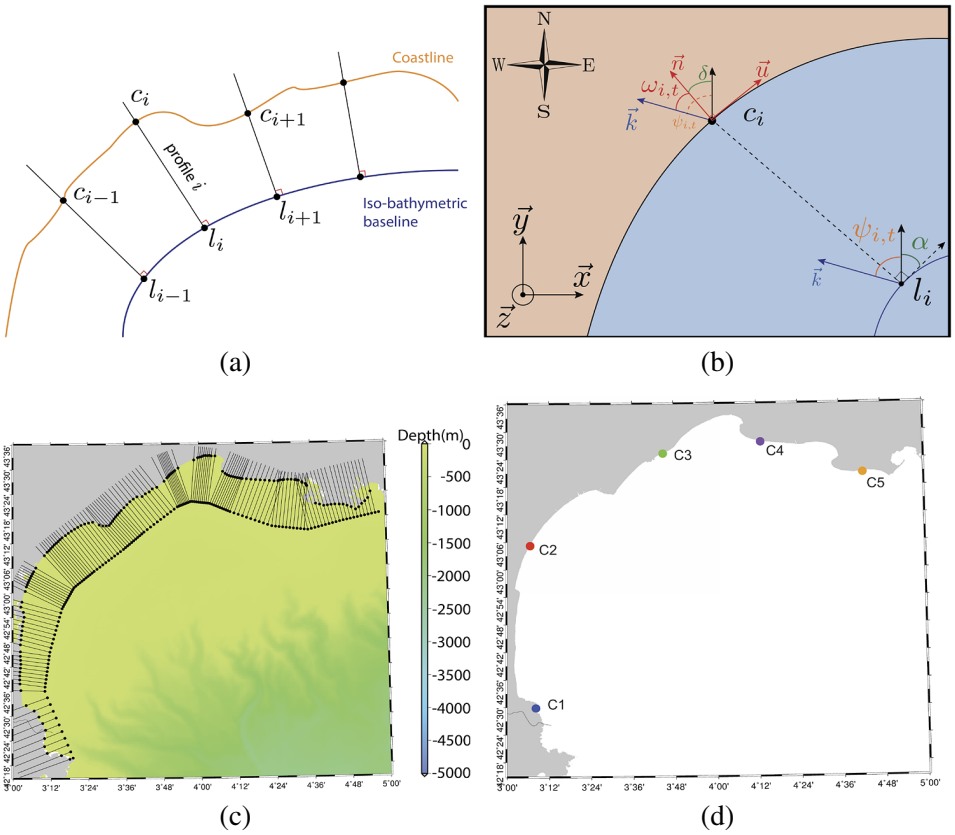


FIG. 7. (a) *A schematic representation of the baseline and the creation of the n profiles*. (b) *Illustration of angles used to compute the impact of the wave energy flux at point* l_i *to its coupled coast point* c_i. ωi_t *denotes the angle of interest*: *the angle between the observed direction of the waves* $\vec{k}$ *at location* l_i*—at a time t—and the cross-shore direction at location* c_i *denoted* $\vec{n}_i$. (c) *The actual profile construction over the GOL. Sea-states conditions are picked-up from a set* $\mathcal{L} = \{l_1, \ldots, l_n\}$ *of n points lying on an iso-bathymetric baseline. From those locations, n profiles normal to the baseline are created. The intersections of those profiles with the coastline derived form a set* $\mathcal{C} = \{c_1, \ldots, c_n\}$ *denoting the reference locations where mass flux energy are derived to. The number n is chosen to fit the resolution required along the shore*. (d) *The selected five locations analysed in the risk analysis*.

energy is classically given by

$$\mathcal{Q}_{i,t} = \frac{1}{8}\rho g \mathrm{Hs}_{i,t}^2 \mathrm{Tp}_{i,t}, \tag{4.2}$$

where ρ denotes the water volumetric mass density and g the gravity constant.

This procedure can be performed both with the storms extracted from the hindcast dataset to monitor the impact of the past events, or with the uplifted storms to assess the impact on the coast of more severe storms.

We compute the long-shore impact at any location c_i for some of the simulated (very) extreme storms. A set of 5 locations from the available c_i [see Figure 7(d)] has been picked as a reference to discuss the assessment of the long-shore impact at the coastline of the GOL under extreme conditions. These locations are manually selected to provide a good covering of the coastline with only few locations for the sake of clarity.

Regarding the angles presented in Figure 7, a positive value of ψ is interpreted as a long-shore contribution in the direction of $\vec{u}$—the tangent at the coast. A negative value is interpreted as a long-shore contribution in the opposite direction, that is, $-\vec{u}$.

Figure 8 gives an overview of the various possibilities offered by the simulation of storms in the assessment of long-shore impact.

First, Figure 8(a) shows the response of the impact model at the 5 reference locations to an uplifted storm at a 100-year return period. Regarding this figure, it is very clear that in this configuration c_2, c_3 and c_4 are impacted towards the west and southwest directions, revealing the presence of an eastern wave forcing. By contrast and since $\psi > 0$ at c_1, this site is impacted towards $\vec{u}$, that is, to the north or northwest at c_1. From such a figure, the time evolution of the long-shore impact regarding the simulated extreme process can be explored.

We may also look at the variability of the long-shore impact when storms vary in extremeness, as defined above. Figure 8(b) represents what could be expected in terms of long-shore impact, at one location and for a given storm uplifted to various return periods.

Another interesting information in the assessment of long-shore impact is to look at the response ψ for several storms uplifted to the 100-year return period. This is illustrated in Figure 8(c) for the point c_5, which is situated at the very east of the GOL. From this figure, we can state that the long-shore impact is likely to be towards the west, catching a consequent amount of energy from the storm coming from the open sea boundary of the GOL (i.e., from the east/southeast). This remark is in accordance with a physical observation that is identified when looking at the shoreline: the formation of sandy spits.

However, and still in Figure 8(c), some of the selected storms have a positive impact during their realisation. This is not really surprising, because as it is located at the edge of the GOL, this shoreline location is also subject to be hit by south

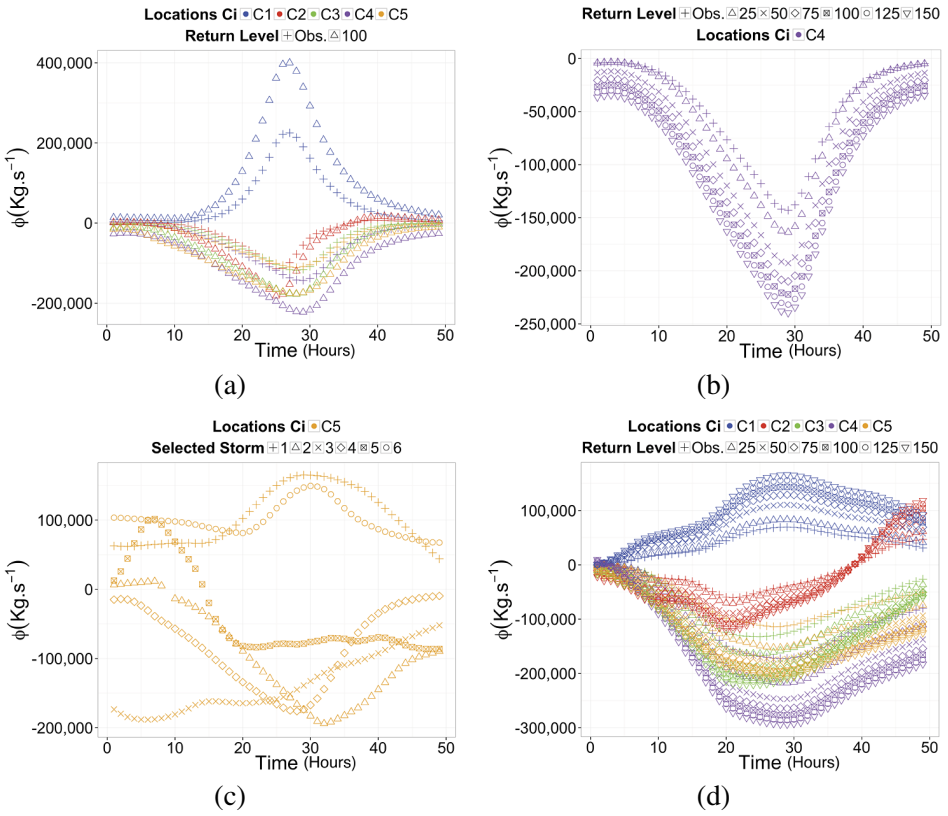


FIG. 8. *Evaluation of the long-shore impact ϕ*; (a) *at the* 5 *locations* c_i *for an observed storm uplifted to the* 100*-year return period*; (b) *at the location* c_4 *for an uplifted storm to the* 25, 50, 75, 100, 125 *and* 150*-year return periods. The impact computed from values of the observed storm are given as well for reference*; (c) *at the location* c_5 *for a sample of observed storms, uplifted to the* 100*-year return period*; (d) *at the* 5 *locations* c_i *for an observed storm uplifted to the* 25, 50, 75, 100, 125 *and* 150*-year return periods. The impact computed from values of the observed storm are also given for reference.*

and southwest storms, which are less frequent but even more damaging than the eastern ones.

Finally, Figure 8(d) is a mix of the possible combinations. It provides a simultaneous preview for various return periods of the storm and at the 5 locations of interest. Spatial patterns of long-shore impact regarding the intensity of a storm might be determined from such a figure.

5. Discussion. We introduced a semiparametric approach to simulate bivariate extreme space–time wave processes. Our motivation was to simulate more extreme storms than those already observed in order to assess event-scale coastal

hazards in such situations. In practice, these storms would feed physical littoral models, which depend strongly on the time evolution of the forcing extreme event.

We applied the methodology presented on a reanalysis dataset covering the GOL area in the northwestern Mediterranean sea.

To demonstrate the benefits of such a method, some simulated storms were used in a risk analysis. Thanks to the simulated processes based on which a control of the extremeness is provided, we showed that the variability of the littoral longshore impact can be assessed, both spatially and through the time evolution. Such results are of the utmost interest in coastal engineering applications, such as the construction of seawalls along the coastline.

This method is especially suitable for its relatively low-cost computational requirement. Indeed, the highest demand concerns the marginal fits, which is an easily parallelisable code. Simulating a set of extreme storms with a physical model would take days where our proposed method will take hours. The proposed method can therefore be applied on massive space–time dataset, as described in this application. Mathematically justified, this method reaches its goal to seamlessly simulate reliable space–time extreme events at a more extreme scale than the ones observed.

However, some limits of the method itself and its implementation should be highlighted. As often when dealing with EVT approaches, we suppose that the underlying dependence structure through time and space is preserved from extreme but observable events to more extreme events. However, it is difficult to physically validate this assumption. As emphasized by Bortot, Coles and Tawn (2000), asymptotic dependence is a limiting property which cannot be verified with certainty from data alone. Usually, the check of the extremal dependence structure relies on modelling properties, arguing from reasonable agreements between empirical and model-based estimates of particular extremal probabilities. Unfortunately, such checking procedures are not possible under our approach, since no particular form of extremal dependence is assumed. As a consequence, if our assumption of a constant space–time extremal dependence for small lags is not satisfied, our approach would lead to an overestimation of the extremal dependence.

Because we are dealing with bivariate space–time processes only, we assume that the third variable defining sea-states conditions [namely the direction $\psi(t)$] remains unchanged in distribution for highest storms. There are good physical reasons to make this hypothesis, such as the GOL orientation, which will never change. Indeed, the open boundary of the GOL, which is southeast oriented, will naturally prohibit the observation of high waves being southeast oriented near the coastline [see Chailan et al. (2014) for further details on the GOL orientation and the implied fetch constraint]. Hence, it seems appropriate to conserve the wave directions from observed storms for heavier storms to keep them physically valid. Consequently, this restriction on the wave directions of the simulated storms to those that have already been observed can be seen both as a strength and a limitation.

From a more practical point of view, it could be argued that the storm size in the Algorithm 1 is fixed and symmetric around the peak value of the storm. This may not reflect the reality for all storms. Therefore, replacing the current fixed size by an adaptive one might be of interest to better represent those storms.

Note that there is no limitation in the methodology to select smaller or longer storms, conditionally to the fact that at least one component excesses its marginal threshold. Regarding coastal risk assessment analysis, selecting smaller storms would result in an underestimation of the length of a storm, and consequently of the overall quantity of interest (e.g., wave energy). By taking storms lasting too long, the opposite may occur. In such cases, selecting the rightful duration of a storm is a true challenge. The use of the extremal coefficient expressing the temporal dependence within storms is by definition a good indicator to determine the storm duration.

In the Algorithm 1, ε is set to avoid the selection of dependent storms and therefore respect statistical assumptions. To avoid dependent storms, it is convenient to always set $\varepsilon = \delta$.

Other parameters of the algorithm can be debated, such as the littoral area S^*. Because its definition is paramount to assess littoral hazards, it could be interesting to evaluate the sensitivity of the storm detection regarding this area.

In this first approach and even if seasonality is found in the data, fixed marginal thresholds are used for the margins transformation. It would be valuable to use more sophisticated expressions of the thresholds to handle the nonstationarity of the data. The use of directional covariates in the thresholds rather than omnidirectional ones might also significantly improve the marginal fits.

In this paper, we have not addressed the estimation uncertainties on marginal fits and their propagation. Block bootstrapping is usually used for assessing such uncertainties. Nevertheless, one practical difficulty is the choice of the blocks to consider, especially in a space–time context. In a similar vein, the validation of the uplifted storms is hard to afford, if not impossible. We recommend using techniques inspired by cross-validation, but practical limitations arise. Uplifted storms are multivariate space–time processes and the first constraint is to find a measure to compare them. Assuming a reduced-dimension measure, the second limit is any uplifted storm that has to be seen as a realisation from the multivariate space–time distribution of storms. Unfortunately, this distribution might only be estimated empirically and many realisations must be used to estimate it correctly. Yet we do not possess enough extreme realisations, by definition. One way to avoid the lack of realisations is to lower the threshold to detect storms in our dataset. Unfortunately, doing so would violate the hypothesis of the method: the need for an exceedance over a "high" threshold, to approximate the asymptotic results.

Beyond those limitations, this method appears promising and opens many perspectives. It would be interesting to extend this approach to the multivariate context because that would allow us to integrate additional variables describing the

environmental phenomenon at a very extreme scale. However, the underlying dependence structures of the considered variables must be thoroughly investigated before being able to justify this extension with the presented assumptions.

Another perspective of work is to apply the method on larger regions. For such an application, the choice of letting $s_{\max}$ be located respectively to each variable should be reviewed. With a wider area, various and independent physical processes might be caught at the peak of storm. Consequently, $\zeta_{i,\mathrm{Hs}}$ and $\zeta_{i,\mathrm{Tp}}$ can be determined from two different processes. On this basis, the uplifted storms might be physically unrealistic.

Other datasets and applications could also be considered. Most notably, we are interested in applying this method in the context of rain-storms. Such an application would allow us to explore the space–time variability of extreme rain-storms scenarios with a plenty set of derived applications. For instance, a simulated scenario can then feed a rainfall-runoff model to study their consequences in terms of floods.

A future work would be the comparison between simulated storms issued by the presented semiparametric approach and those issued by other parametric approaches, and in particular the generalized Pareto processes. Such a comparison would be valuable since both approaches present similarities.

At the same time and after having performed a small risk analysis using some of the simulated extreme space–time waves events, one challenge is to use those storms to feed heavy computational physical models assessing other coastal hazards, such as a flood overland model. In our opinion, this challenge may represent the foundation of the next generation of coastal flood early warning systems such as Delaware's Coastal Flood Monitoring System (CFMS).[2]

APPENDIX SECTION

A mathematical justification of the storm uplift can be obtained through the following asymptotic equivalence for conditional distributions.

Indeed, following Caires, de Haan and Smith (2011),

$$P\left(\frac{T^{\leftarrow}(\zeta_i Z_i^*(s,t)) - b_{n\zeta_i}}{a_{n\zeta_i}} \in A \Big| \max_{s\in\mathcal{M}^*} Z_i^*(s,t) > 1\right)$$

has the same limit (as $n \to \infty$) as

$$P\left(\frac{Z_i(s,t) - b_n}{a_n} \in A \Big| \max_{s\in\mathcal{M}^*} Z_i^*(s,t) > 1\right),$$

where $Z_i^*(s,t) = [1 + \xi(s,t)(Z_i(s,t) - b_n(s,t))/a_n(s,t)]^{1/\xi(s,t)}$ and $T^{\leftarrow}(y) = b_n + a_n\frac{y^{\xi}-1}{\xi}$.

[2] http://coastal-flood.udel.edu/

Let us drop both i and (s,t) indexes for the sake of simplicity in the left part of the conditional probability. The former limit equivalence is valid since following Ferreira and de Haan (2014)-Section 4.2,

$$
\begin{aligned}
&P\left(\frac{T^{\leftarrow}(\zeta Z^*)-b_{n\zeta}}{a_{n\zeta}}\in A\Big|\max_{s\in\mathcal{M}^*}Z_i^*(s,t)>1\right)\\
&\quad=P\left(\frac{a_n}{a_{n\zeta}}\frac{[\zeta^{\xi}(1+\xi\frac{Z-b_n}{a_n})]-1}{\xi}-\frac{b_{n\zeta}-b_n}{a_{n\zeta}}\in A\Big|\max_{s\in\mathcal{M}^*}Z_i^*(s,t)>1\right)\\
&\quad=P\left(\frac{a_n\zeta^{\xi}}{a_{n\zeta}}\frac{1+\xi\frac{Z-b_n}{a_n}-\zeta^{-\xi}}{\xi}-\frac{b_{n\zeta}-b_n}{a_{n\zeta}}\in A\Big|\max_{s\in\mathcal{M}^*}Z_i^*(s,t)>1\right)\\
&\quad=P\left(\frac{a_n\zeta^{\xi}}{a_{n\zeta}}\left(\frac{Z-b_n}{a_n}-\zeta^{-\xi}\left[\frac{b_{n\zeta}-b_n}{a_n}-\frac{\zeta^{\xi}-1}{\xi}\right]\right)\in A\Big|\max_{s\in\mathcal{M}^*}Z_i^*(s,t)>1\right)\\
&\quad=P\left(\frac{Z-b_n}{a_n}\in\frac{a_{n\zeta}\zeta^{-\xi}}{a_n}A+\zeta^{-\xi}\left(\frac{b_{n\zeta}-b_n}{a_n}-\frac{\zeta^{\xi}-1}{\xi}\right)\Big|\max_{s\in\mathcal{M}^*}Z_i^*(s,t)>1\right).
\end{aligned}
$$

From de Haan and Ferreira (2006), it can be deduced that:

1. $\lim_{n\to\infty}\frac{a_{n\zeta}}{a_n}=\zeta^{\xi}$ (see proof of Lemma 1.2.9, p. 24);
2. $\lim_{n\to\infty}\frac{b_{n\zeta}-b_n}{a_n}=\frac{\zeta^{\xi}-1}{\xi}$ [consider $U(n)$ as in Theorem 1.1.2 for decomposing $\frac{b_{n\zeta}-b_n}{a_n}$ as $\frac{U(n)-b_n}{a_n}-\frac{U_{n\zeta}-b_{n\zeta}}{a_{n\zeta}}\frac{a_{n\zeta}}{a_n}+\frac{U(n\zeta)-U(n)}{a_n}$ and use 1.1.20, p. 10].

Then

$$
\frac{a_{n\zeta}\zeta^{-\xi}}{a_n}\to 1\quad\text{and}\quad\zeta^{-\xi}\left(\frac{b_{n\zeta}-b_n}{a_n}-\frac{\zeta^{\xi}-1}{\xi}\right)\to 0
$$

uniformly for $(s,t)\in S\times\mathcal{T}$ as $n\to\infty$ and the result follows using the convergence to types theorem [see Embrechts, Klüppelberg and Mikosch (1997), Theorem A1.5].

There is no limitation to extend this reasoning to a bivariate context. In accordance with our chosen approach, only one of the two considered processes is concerned with the conditional event. Indeed, the conditional event has no impact on the aforementioned probability developments. Hence, we can similarly show that

$$
\begin{aligned}
P\Bigg(&\frac{T_1^{\leftarrow}(\zeta_{1,i}Z_{1,i}^*(s,t))-b_{1,n\zeta_i}}{a_{1,n\zeta_i}}\in A_1,\\
&\frac{T_2^{\leftarrow}(\zeta_{2,i}Z_{2,i}^*(s,t))-b_{2,n\zeta_i}}{a_{2,n\zeta_i}}\in A_2\Big|\max_{s\in\mathcal{M}^*}Z_{1,i}^*(s,t)>1\Bigg),
\end{aligned}
$$

where $T_1^{\leftarrow}(y) = b_{1,n} + a_{1,n}\frac{y^{\xi_1}-1}{\xi_1}$ and $T_2^{\leftarrow}(y) = b_{2,n} + a_{2,n}\frac{y^{\xi_2}-1}{\xi_2}$, has the same limit (as $n \to \infty$) as

$$P\Big(\frac{Z_{1,i}(s,t) - b_{1,n}}{a_{1,n}} \in A_1, \frac{Z_{2,i}(s,t) - b_{2,n}}{a_{2,n}} \in A_2 \Big| \max_{s\in\mathcal{M}^*} Z_{1,i}^*(s,t) > 1\Big).$$

Acknowledgments. We would like to thank A. Laurent and F. Bouchette who have both contributed significantly to this document.

This work was performed within the context of both the LITTOCMS and MISTRAL-LITTORAL projects. It also forms part of the GLADYS initiative; see www.gladys-littoral.org.

REFERENCES

BAGNOLD, R. A. (1966). An approach to the sediment transport problem. General Physics Geological Survey, Prof. Paper.

BECHLER, A., BEL, L. and VRAC, M. (2015). Conditional simulations of the extremal t process: Application to fields of extreme precipitation. *Spat. Stat.* **12** 109–127. MR3346645

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes. Theory and Applications*. Wiley, Chichester. MR2108013

BORTOT, P., COLES, S. and TAWN, J. (2000). The multivariate Gaussian tail model: An application to oceanographic data. *J. Roy. Statist. Soc. Ser. C* **49** 31–49. MR1817873

BOUCHETTE, F., SABATIER, F., SYLAIOS, G., MEULÉ, S., LIOU, J. L., HEURTEFEUX, H., DENAMIEL, C. and HWUNG, W. (2012). SUBDUNE tool: Quasiexplicit formulation of the water level along the shoreline. *Rev. Paralia* **12** 223–232.

BRUNEL, C., CERTAIN, R., SABATIER, F., ROBIN, N., BARUSSEAU, J. P., ALEMAN, N. and RAYNAL, O. (2014). 20th century sediment budget trends on the Western Gulf of Lions shoreface (France): An application of an integrated method for the study of sediment coastal reservoirs. *Geomorphology* **204** 625–637.

CAIRES, S., DE HAAN, L. and SMITH, R. L. (2011). On the determination of the temporal and spatial evolution of extreme events Technical Report, Deltares. Report 1202120-001-HYE-004 (for Rijkswaterstaat, Centre for Water Management).

CAMPMAS, L., BOUCHETTE, F., MEULE, S., PETITJEAN, L., SOUS, D., LIOU, J.-Y., LEROUX-MALLOUF, R., SABATIER, F. and HWUNG, H.-H. (2014). Typhoons driven morphodynamics of the Wan Tzu Liao sand barrier (Taiwan). *Coastal Eng. Proc.* **1** sediment–50.

CASTRUCCIO, S., HUSER, R. and GENTON, M. G. (2016). High-order composite likelihood inference for max-stable distributions and processes. *J. Comput. Graph. Statist.* **25** 1212–1229. MR3572037

CERC (1984). Shore Protection Manual.

CHAILAN, R. (2015). Application of Scientific Computing and Statistical Analysis to Address Coastal Hazards Ph.D. thesis University of Montpellier.

CHAILAN, R., TOULEMONDE, G., BOUCHETTE, F., LAURENT, A., SEVAULT, F. and MICHAUD, H. (2014). Spatial assessment of extreme significant waves heights in the Gulf of Lions. *Coastal Eng. Proc.* **1** management–17.

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London. MR1932132

DAVIS, R. A., KLÜPPELBERG, C. and STEINKOHL, C. (2013a). Max-stable processes for modeling extremes observed in space and time. *J. Korean Statist. Soc.* **42** 399–414. MR3255398

DAVIS, R. A., KLÜPPELBERG, C. and STEINKOHL, C. (2013b). Statistical inference for max-stable processes in space and time. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 791–819. MR3124792

DAVISON, A. C. and HUSER, R. (2015). Statistics of extremes. *Annu. Rev. Stat. Appl.* **2** 203–235.

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 161–186. MR2963980

DE HAAN, L. (1984). A spectral representation for max-stable processes. *Ann. Probab.* **12** 1194–1204. MR0757776

DE HAAN, L. and DE RONDE, J. (1998). Sea and wind: Multivariate extremes at work. *Extremes* **1** 7–45. MR1652944

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*. Springer, New York. MR2234156

DE HAAN, L. and LIN, T. (2001). On convergence toward an extreme value distribution in $C[0, 1]$. *Ann. Probab.* **29** 467–483. MR1825160

DIEKER, A. B. and MIKOSCH, T. (2015). Exact simulation of Brown-Resnick random fields at a finite number of locations. *Extremes* **18** 301–314. MR3351818

DOMBRY, C., ENGELKE, S. and OESTING, M. (2016). Exact simulation of max-stable processes. *Biometrika* **103** 303–317. MR3509888

DOMBRY, C. and EYI-MINKO, F. (2013). Regular conditional distributions of continuous max-infinitely divisible random fields. *Electron. J. Probab.* **18** 1–21. MR3024101

DOMBRY, C., ÉYI-MINKO, F. and RIBATET, M. (2013). Conditional simulation of max-stable processes. *Biometrika* **100** 111–124. MR3034327

DOMBRY, C. and RIBATET, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Stat. Interface* **8** 9–17. MR3320385

EASTOE, E. F. and TAWN, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58** 25–45. MR2662232

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events. Applications of Mathematics* (*New York*) **33**. Springer, Berlin. MR1458613

EMBRECHTS, P., KOCH, E. and ROBERT, C. (2016). Space-time max-stable models with spectral separability. *Adv. in Appl. Probab.* **48** 77–97. MR3539298

ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. and SCHLATHER, M. (2015). Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 239–265. MR3299407

FERREIRA, A. and DE HAAN, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20** 1717–1737. MR3263087

GOULDBY, B., MÉNDEZ, F. J., GUANCHE, Y., RUEDA, A. and MÍNGUEZ, R. (2014). A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis. *Coastal Eng.* **88** 15–26.

GROENEWEG, J., CAIRES, S. and ROSCOE, K. (2012). Temporal and spatial evolution of extreme events. *Coastal Eng. Proc.* **1** management–9.

GUTIERREZ, B. T., PLANT, N. G., THIELER, E. R. and TURECEK, A. (2015). Using a Bayesian network to predict barrier island geomorphologic characteristics. *J. Geophys. Res., Earth Surf.* **120** 2452–2475.

HERRMANN, M. and SOMOT, S. (2008). Relevance of ERA40 dynamical downscaling for modeling deep convection in the Mediterranean Sea. *Geophys. Res. Lett.* **35**.

HERRMANN, M., SEVAULT, F., BEUVIER, J. and SOMOT, S. (2010). What induced the exceptional 2005 convection event in the northwestern Mediterranean basin? Answers from a modeling study. *J. Geophys. Res., Oceans* (1978–2012) **115**.

HUSER, R. and DAVISON, A. C. (2013). Composite likelihood estimation for the Brown-Resnick process. *Biometrika* **100** 511–518. MR3068451

HUSER, R. and DAVISON, A. C. (2014). Space-time modelling of extreme events. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 439–461. MR3164873

JONATHAN, P., EWANS, K. and RANDELL, D. (2013). Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Eng*. **79** 22–31.

LANTUÉJOUL, C. and BEL, L. (2014). Simulation conditionnelle du processus de Schlather. In 46*èmes Journées de Statistique de la SFdS*.

MICHAUD, H. (2011). Impacts des vagues sur les courants marins: Modélisation multi-échelle de la plage au plateau continental Ph.D. thesis Université Montpellier II-Sciences et Techniques du Languedoc.

MICHAUD, H., ROBIN, N., ESTOURNEL, C., MARSALEIX, P., LEREDDE, Y., CERTAIN, R. and BOUCHETTE, F. (2013). 3D hydrodynamic modeling of a microtidal barred beach (Sète, NW Mediterranean Sea) during storm conditions. In *Proc*. 7*th Int*. *Conf*. *on Coastal Dynamics*, *Arcachon France* **139** 1183–1194.

PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J*. *Amer*. *Statist*. *Assoc*. **105** 263–277. MR2757202

RAILLARD, N., AILLIOT, P. and YAO, J. (2014). Modeling extreme values of processes observed at irregular time steps: Application to significant wave height. *Ann*. *Appl*. *Stat*. **8** 622–647. MR3192005

RASCLE, N. and ARDHUIN, F. (2013). A global wave parameter database for geophysical applications. Part 2: Model validation with improved source term parameterization. *Ocean Model*. **70** 174–188.

RIBATET, M., COOLEY, D. and DAVISON, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statist*. *Sinica* **22** 813–845. MR2954363

SCHLATHER, M. and TAWN, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90** 139–156. MR1966556

SHABY, B. A. (2014). The open-faced sandwich adjustment for MCMC using estimating functions. *J*. *Comput*. *Graph*. *Statist*. **23** 853–876. MR3224659

SHABY, B. A. and REICH, B. J. (2012). Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics* **23** 638–648. MR3019056

SMITH, R. L. (1990). Max-stable processes and spatial extremes. Preprint. Univ. Surrey.

THIBAUD, E. and OPITZ, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika* **102** 855–870. MR3431558

TOLMAN, H. L. (2014). User Manual and System Documentation of WAVEWATCH III® version 4.18. Technical Report 316.

WADSWORTH, J. L. and TAWN, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* **101** 1–15. MR3180654

WANG, Y. and STOEV, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Adv*. *in Appl*. *Probab*. **43** 461–483. MR2848386

UNIVERSITY OF MONTPELLIER
2 PLACE EUGÈNE BATAILLON
34095 MONTPELLIER CEDEX 5
OCCITANIE
FRANCE
E-MAIL: romain.chailan@umontpellier.fr
gwladys.toulemonde@umontpellier.fr
jean-noel.bacro@umontpellier.fr

Annexe C

G3 - Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. JASA (2019).

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
https://doi.org/10.1080/01621459.2019.1617152

Hierarchical Space-Time Modeling of Asymptotically Independent Exceedances With an Application to Precipitation Data

Jean-Noël Bacro[a], Carlo Gaetan[b], Thomas Opitz[c], and Gwladys Toulemonde[d]

[a]IMAG, Université de Montpellier, CNRS, Montpellier, France; [b]DAIS, Università Ca' Foscari di Venezia, Venice, Italy; [c]BioSP, INRA, Avignon, France; [d]IMAG, Univ Montpellier, CNRS, INRIA, Montpellier, France

ABSTRACT

The statistical modeling of space-time extremes in environmental applications is key to understanding complex dependence structures in original event data and to generating realistic scenarios for impact models. In this context of high-dimensional data, we propose a novel hierarchical model for high threshold exceedances defined over continuous space and time by embedding a space-time Gamma process convolution for the rate of an exponential variable, leading to asymptotic independence in space and time. Its physically motivated anisotropic dependence structure is based on geometric objects moving through space-time according to a velocity vector. We demonstrate that inference based on weighted pairwise likelihood is fast and accurate. The usefulness of our model is illustrated by an application to hourly precipitation data from a study region in Southern France, where it clearly improves on an alternative censored Gaussian space-time random field model. While classical limit models based on threshold-stability fail to appropriately capture relatively fast joint tail decay rates between asymptotic dependence and classical independence, strong empirical evidence from our application and other recent case studies motivates the use of more realistic asymptotic independence models such as ours. Supplementary materials for this article, including a standardized description of the materials available for reproducing the work, are available as an online supplement.

ARTICLE HISTORY
Received August 2017
Revised February 2019

KEYWORDS
Asymptotic independence;
Composite likelihood;
Gamma random fields;
Hourly precipitation;
Space-time convolution;
Space-time extremes.

1. Introduction

The French Mediterranean area is subject to heavy rainfall events occurring mainly in the fall season. Intense precipitation often leads to flash floods, which can be defined as a sudden strong rise of the water level. Flash floods often cause fatalities and important material damage. In the literature, such intense rainfalls are often called flood-risk rainfall (Carreau and Bouvier 2016); characterizing their spatio-temporal dependencies is key to understanding these processes. In this article, we consider a large dataset of hourly precipitation measurements from a study region in Southern France. We tackle the challenge of proposing a physically interpretable statistical space-time model for high threshold exceedances, which aims to capture the complex dependence and time dynamics of the data process.

Fueled by important environmental applications during the last decade, the statistical modeling of spatial extremes has undergone a fast evolution. A shift from maxima-based modeling to approaches using threshold exceedances can be observed over recent years, whose reasons lie in the capability of thresholding techniques to exploit more information from the data and to explicitly model the original event data. A first overview of approaches to modeling maxima is due to Davison, Padoan, and Ribatet (2012). A number of hierarchical models based on latent Gaussian processes (Casson and Coles 1999; Gaetan and Grigoletto 2007; Cooley, Nychka, and Naveau 2007; Sang and Gelfand 2009) have been proposed, but they may be criticized for relying on the rather rigid Gaussian dependence with very weak dependence in the tail, while the lack of closed-form marginal distributions makes interpretation difficult and frequentist inference cumbersome. Max-stable random fields (Smith 1990; Schlather 2002; Kabluchko, Schlather, and de Haan 2009; Davison and Gholamrezaee 2012; Reich and Shaby 2012; Opitz 2013) are the natural limit models for maxima data and have spawned a very rich literature, from which the model of Reich and Shaby (2012) stands out for its hierarchical construction simplifying high-dimensional multivariate calculations and Bayesian inference. Generalized Pareto (GP) processes (Ferreira and de Haan 2014; Opitz, Bacro, and Ribereau 2015; Thibaud and Opitz 2015) are the equivalent limit models for threshold exceedances. However, the asymptotic dependence stability in these limiting processes for maxima and threshold exceedances has a tendency to be overly restrictive when asymptotic dependence strength decreases at high levels and may vanish ultimately in the case of asymptotic independence. The results from the empirical spatio-temporal exploration of our French rainfall data in Section 6.2 are strongly in favor of asymptotic independence, which appears to be characteristic for many environmental datasets (Davison, Huser, and Thibaud 2013; Thibaud, Mutzner, and Davison 2013; Tawn et al. 2018) and may arise from physical laws such as the conservation of mass. This

CONTACT Carlo Gaetan gaetan@unive.it DAIS, Università Ca' Foscari di Venezia, Venice 30123, Italy.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JASA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.
These materials were reviewed for reproducibility.

has motivated the development of more flexible dependence models, such as max-mixtures of max-stable and asymptotically independent processes (Wadsworth and Tawn 2012; Bacro, Gaetan, and Toulemonde 2016) or max-infinitely divisible constructions (Huser, Opitz, and Thibaud 2018) for maxima data, or Gaussian scale mixture processes (Opitz 2016; Huser, Opitz, and Thibaud 2017) for threshold exceedances, capable to accommodate asymptotic dependence, asymptotic independence and Gaussian dependence with a smooth transition. Other flexible spatial constructions involve marginally transformed Gaussian processes (Huser and Wadsworth 2019). Such threshold models can be viewed as part of the wider class of copula models (see Bortot, Coles, and Tawn 2000; Davison, Huser, and Thibaud 2013, for other examples) typically combining univariate limit distributions with dependence structures that should ideally be flexible and relatively easy to handle in practice.

Statistical inference is then often carried out assuming temporal independence in measurements typically observed at spatial sites at regularly spaced time intervals. However, developing flexible space-time modeling for extremes is crucial for characterizing the temporal persistence of extreme events spanning several time steps; such models are important for short-term prediction in applications such as the forecasting of wind power and atmospheric pollution, and for extreme event scenario generators providing inputs to impact models, for instance in hydrology and agriculture. Early spatio-temporal models for rainfall were proposed in the 1980s (Rodriguez-Iturbe, Cox, and Isham 1987; Cox and Isham 1988, and the references therein) and exploit the idea that storm events give rise to a cluster of rain cells, which are represented as cylinders in space-time. Currently, only few statistical space-time models for extremes are available. Davis and Mikosch (2008) considered extremal properties of heavy-tailed moving average processes where coefficients and the white-noise process depend on space and time, but their work was not focused on practical modeling. Sang and Gelfand (2009) proposed a hierarchical procedure for maxima data but limited to latent Gauss–Markov random fields. Davis, Klüppelberg, and Steinkohl (2013a, 2013b) extended the widely used class of Brown–Resnick max-stable processes to the space-time framework and propose pairwise likelihood inference. Spatial max-stable processes with random set elements have been proposed by Schlather (2002) and Davison and Gholamrezaee (2012), and Huser and Davison (2014) have fitted a space-time version to threshold exceedances of hourly rainfall data through pairwise censored likelihood. Huser and Davison (2014) modeled storms as discs of random radius moving at a random velocity for a random duration, leading to randomly centered space-time cylinders; our models developed in the following rely on similar geometric representations. A Bayesian approach based on spatial skew-t random fields with a random set element and temporal autoregression was proposed by Morris et al. (2017). The aforementioned space-time models may capture asymptotic dependence or exact independence at small distances but are unsuitable for dealing with residual dependence in asymptotic independence. In this article, we propose a novel approach to space-time modeling of asymptotically independent data to avoid the tendency of max-stable-like models to potentially strongly overestimate joint extreme risks. In a similar context, Nieto-Barajas and Huerta (2017) recently proposed a spatio-temporal Pareto model for heavy-tailed data on spatial lattices, generalizing the temporal latent process model of Bortot and Gaetan (2014) to space-time.

Our model provides a hierarchical formulation for modeling spatio-temporal exceedances over high thresholds. It is defined over a continuous space-time domain and allows for a physical interpretation of extreme events spreading over space and time. Strong motivation also comes from Bortot and Gaetan (2014) by developing a generalization of their latent temporal process. Alternatively, our latent process could be viewed as a space-time version of the temporal trawl processes introduced by Barndorff-Nielsen et al. (2014) and exploited for extreme values by Noven, Veraart, and Gandy (2015), with spatial extensions recently proposed by Opitz (2017). Our approach is based on representing a GP distribution as a Gamma mixture of an exponential distribution, enabling us to keep easily tractable marginal distributions which remain coherent with univariate extreme value theory. We use a kernel convolution of a space-time Gamma random process (Wolpert and Ickstadt 1998a) based on influence zones defined as cylinders with an ellipsoidal basis to generate anisotropic spatio-temporal dependence in exceedances. We then develop statistical inference based on pairwise likelihood.

The article is structured as follows. Our hierarchical model with a detailed description of its two stages and marginal transformations is developed in Section 2. Space-time Gamma random fields are presented in Section 2.1 where we also propose the construction and formulas for the space-time objects used for kernel convolution. Section 3 characterizes tail dependence behavior in our new model yielding asymptotic independence in space and time. Statistical inference of model parameters is addressed in Section 4 based on a pairwise log-likelihood for the observed censored excesses. We show good estimation performance through a simulation study presented in Section 5 involving two scenarios of different complexity. In Section 6, we focus on the dataset and explore in detail how our fitted space-time model captures spatio-temporal extremal dependence in hourly precipitation. Since a natural choice of a reference model for asymptotically independent data is to use threshold-censored space-time Gaussian processes, we show the good relative performance of our model by comparing it to such alternatives. A discussion of our modeling approach with some potential future extensions closes the article in Section 7.

2. A Hierarchical Model for Spatio-Temporal Exceedances

When dealing with exceedances of a random variable X above a high threshold u, univariate extreme value theory suggests using the limit distribution of GP type. The GP cumulative distribution function (cdf) is defined for any $y > 0$ by

$$\mathrm{GP}(y;\sigma,\xi) = 1 - \left(1 + \xi\frac{y}{\sigma}\right)_+^{-(1/\xi)}, \qquad (1)$$

where $(a)_+ = \max(0, a)$, ξ is a shape parameter and σ a positive scale parameter. The sign of ξ characterizes the maximum domain of attraction of the distribution of X: $\xi > 0$ corresponds

to the Fréchet domain of attraction while $\xi = 0$ and $\xi < 0$ correspond to the Gumbel and Weibull ones, respectively.

When $\xi > 0$, the GP distribution can be expressed as a Gamma mixture for the rate of the exponential distribution (Reiss and Thomas 2007, p. 157), that is,

$$V|\Lambda \sim \text{Exp}(\Lambda), \quad \Lambda \sim \text{Gamma}(1/\xi, \sigma/\xi) \quad \Rightarrow \quad V \sim \text{GP}(\cdot; \sigma, \xi), \tag{2}$$

where $\text{Exp}(b)$ refers to the Exponential distribution with rate $b > 0$ and $\text{Gamma}(a, b)$ to the Gamma distribution with shape $a > 0$ and rate $b > 0$. Based on this hierarchical structure, we will here develop a stationary space-time construction for modeling exceedances over a high threshold, which possesses marginal GP distributions for the strictly positive excess above the threshold.

2.1. First Stage: Generic Hierarchical Space-Time Structure

We consider a stationary space-time random field $Z = \{Z(x), x \in \mathcal{X}\}$ with $x = (s, t)$ and $\mathcal{X} = \mathbb{R}^2 \times \mathbb{R}^+$, such that s indicates spatial location and t time. Without loss of generality, we assume that the margins $Z(x)$ belong to the Fréchet domain of attraction with positive shape parameter ξ. To infer the tail behavior of $\{Z(x)\}$, we focus on values exceeding a fixed high threshold u, and we consider the exceedances over u,

$$Y(x) = (Z(x) - u) \cdot \mathbf{1}_{(u,\infty)}(Z(x)). \tag{3}$$

Standard results from extreme value theory (de Haan and Ferreira 2006) establish the *GP* distribution with $\xi > 0$ in (1) as the limit of suitably renormalized positive threshold exceedances in (3), such that it represents a natural model for the values $Y(x) > 0$. Following Bortot and Gaetan (2014), we use the representation of the GP distribution as a Gamma mixture of an exponential distribution to formulate a two-stage model that induces spatio-temporal dependence arising in both the exceedance indicators $\mathbf{1}_{(u,\infty)}(Z(x))$ and the positive excesses $Z(x) - u > 0$ by integrating space-time dependence in a latent Gamma component. A key feature of our model is that it naturally links exceedance probability to the size of the excess and therefore provides a joint space-time structure of the zero part and the positive part in the zero-inflated distribution of $Y(x)$.

In the first stage, we condition on a latent space-time random field $\{\Lambda(x)\}$ with marginal distributions $\Lambda(x) \sim \text{Gamma}(\alpha, \beta)$ and assume that

$$Y(x) \mid [\Lambda(x), Y(x) > 0] \sim \text{Exp}(\Lambda(x)), \tag{4a}$$

$$\Pr(Y(x) > 0 \mid \Lambda(x)) = e^{-\kappa \Lambda(x)}, \tag{4b}$$

where $\kappa > 0$ is a parameter controlling the rate of upcrossings of the threshold. The resulting marginal distribution of $Y(x)$ conditionally on $Z(x) > u$ corresponds to the GP distribution, and the unconditional marginal cdf of $Y(x)$ is

$$F(y; \sigma, \xi) = \begin{cases} p & \text{for } y = 0, \\ p + (1-p)\text{GP}(y; \xi, \sigma) & \text{for } y > 0, \end{cases} \tag{5}$$

with shape parameter $\xi = 1/\alpha$, scale parameter $\sigma = (\kappa + \beta)/\alpha$, and with $1 - p$ the probability of an exceedance over u, that is, $\Pr(Z(x) > u) = \Pr(Y(x) > 0) = 1 - p$. The probability of exceeding u,

$$\Pr(Z(x) > u) = \text{E}\left(\Pr(Y(x) > 0 | \Lambda(x))\right) = \text{E}\left(e^{-\kappa \Lambda(x)}\right) = \left(\frac{\beta}{\kappa + \beta}\right)^{\alpha} \tag{6}$$

depends on κ and corresponds to the Laplace transform of $\Lambda(x)$ evaluated at κ. The constraint $\xi > 0$ is not restrictive for dealing with precipitation in the French Mediterranean area, which is known to be heavy-tailed. For general modeling purposes, we can relax this assumption by following Bortot and Gaetan (2016): we consider a marginal transformation within the class of GP distributions for threshold exceedances, for which we suppose that $\alpha = 1$ and $\beta = 1$ for identifiability. By transforming $Y(x)$ through the probability integral transform

$$g(y) = \text{GP}^{-1}(\text{GP}(y; 1, 1 + \kappa); \sigma^*, \xi^*) \tag{7}$$

$$= (\sigma^*/\xi^*)\left\{\left(1 + \frac{y}{\kappa + 1}\right)^{\xi^*} - 1\right\} \tag{8}$$

with parameters $\xi^* \in \mathbb{R}$ and $\sigma^* > 0$ to be estimated, we get a marginally transformed random field $Y^*(x) = g(Y(x))$ which satisfies $Y^*(x) \sim \text{GP}(\cdot; \xi^*, \sigma^*)$, conditionally on $Y^*(x) > 0$. Notice that it is straightforward to develop extensions with nonstationary marginal excess distributions by injecting response surfaces $\sigma^\star(x)$ and $\xi^\star(x)$ into (8). Moreover, nonstationarity could be introduced into the latent Gamma model (4) in different ways. If $\kappa = \kappa(x)$ depends on x or other covariate information, exceedance probabilities become nonstationary. If Gamma parameters $\alpha = \alpha(x)$ and $\beta = \beta(x)$ depend on covariates, then the GP margins in $Y(x)$ become nonstationary. Finally, one could combine the two previous nonstationary extensions.

2.2. Second Stage: Space-Time Dependence With Gamma Random Fields

Spatio-temporal dependence is introduced by means of a space-time stationary random field $\{\Lambda(x), x \in \mathcal{X}\}$ with $\text{Gamma}(\alpha, \beta)$ marginal distributions. In principle, we could use an arbitrarily wide range of models with any kind of space-time dependence structure, for instance by marginally transforming a space-time Gaussian random field using the copula idea (Joe 1997). However, we here aim to propose a construction where Gamma marginal distributions arise naturally without applying rather artificial marginal transformations. Inspired by the Gamma process convolutions of Wolpert and Ickstadt (1998a), we develop a space-time Gamma convolution process with Gamma marginal distributions. The kernel shape in our construction allows for a straightforward interpretation of the dependence structure, and it offers a physical interpretation of real phenomena such as mass and particle transport. Moreover, we obtain simple analytical formulas for the bivariate distributions, which facilitates statistical inference, interpretation and the characterization of joint tail properties.

We fix $\mathcal{X} = \mathbb{R}^3$ and consider $A \in \mathcal{B}_b(\mathcal{X})$, a subset of $\mathcal{X}$ belonging to the σ-field $\mathcal{B}_b(\mathcal{X})$ restricted to bounded sets of $\mathcal{X}$. A Gamma random field $\Gamma(dx)$ (Ferguson 1973) is a nonnegative random measure defined on $\mathcal{X}$ characterized by a base measure $\alpha(dx)$ and a rate parameter β such that

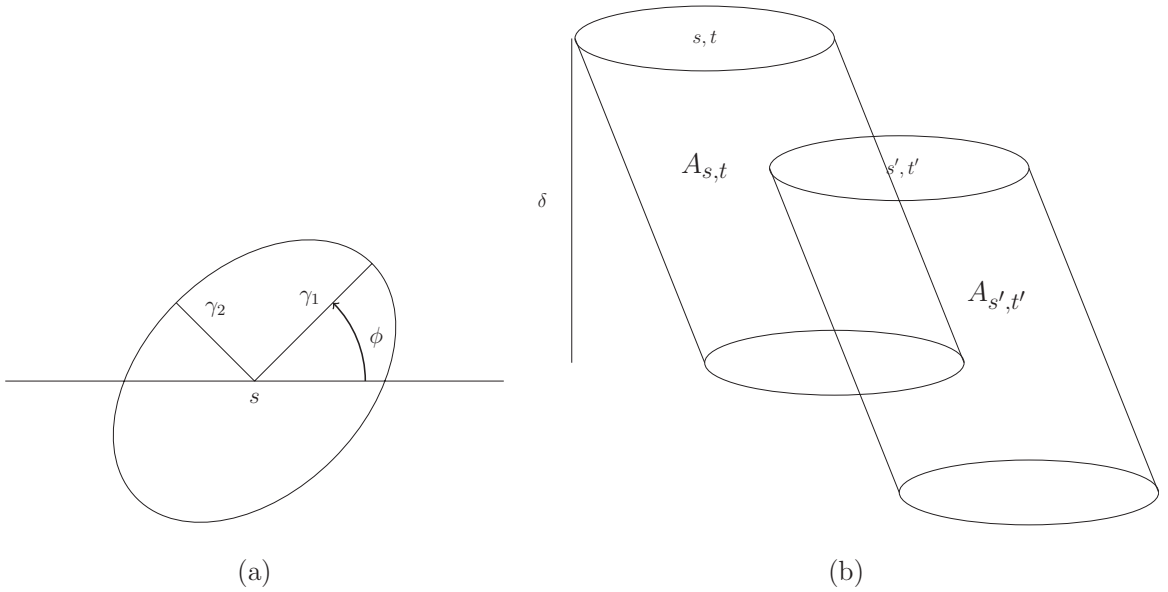


Figure 1. Space-time kernels. Left display: a spatial ellipse $E(s,\gamma_1,\gamma_2,\phi)$ centered at s. Right display: an intersection of two slated elliptical cylinders $A_{s,t}$ and $A_{s',t'}$ with duration δ.

1. $\Gamma(A) := \int_A \Gamma(dx) \sim \text{Gamma}(\alpha(A),\beta)$, with $\alpha(A) := \int_A \alpha(dx)$;
2. for any $A_1, A_2 \in \mathcal{B}_b(\mathcal{X})$ such that $A_1 \cap A_2 = \emptyset$, $\Gamma(A_1)$ and $\Gamma(A_2)$ are independent random variables.

The calculation of important formulas in this article requires the Laplace exponent of the random measure given as

$$\begin{aligned}\mathcal{L}(\phi) &:= -\log \mathrm{E}\left(\exp\left\{-\int \phi(x)\Gamma(dx)\right\}\right)\\ &= \int_{\mathcal{X}} \log\left\{1+\frac{\phi(x)}{\beta}\right\}\alpha(dx),\end{aligned}$$

where ϕ is any positive measurable function; in our case, it will represent the kernel function (see Appendix A). We propose to model $\{\Lambda(x), x \in \mathcal{X}\}$ as a convolution using a three-dimensional indicator kernel $K(x,x')$ with an indicator set of finite volume used to convolve the Gamma random field $\Gamma(dx)$ (Wolpert and Ickstadt 1998a), that is, $\Lambda(x) = \int K(x,x')\Gamma(dx')$. The shape of the kernel can be very general (although nonindicator kernels usually do not lead to Gamma marginal distributions), and particular choices may lead to nonstationary random fields, or to stationary random fields with given dependence properties such as full symmetry, separability or independence beyond some spatial distance or temporal lag. To limit model complexity and computational burden to a reasonable amount, we use the indicator kernel $K(x,x') = \mathbf{1}_A(x-x')$, for $A \in \mathcal{B}_b(\mathcal{X})$, where A is given as a slated elliptical cylinder, defining a three-dimensional set A_x that moves through $\mathcal{X}$ according to some velocity vector. More precisely, let $E(s,\gamma_1,\gamma_2,\phi)$ be an ellipse centered at $s = (s_1,s_2) \in \mathbb{R}^2$ (see Figure 1(a)), whose axes are rotated counterclockwise by the angle ϕ with respect to the coordinate axes, whose semi-axes' lengths in the rotated coordinate system are γ_1 and γ_2, respectively. A physical interpretation is that the ellipse describes the spatial influence zone of a storm centered at s. For the temporal dynamics, we assume that the ellipses (storms) $E(s,\gamma_1,\gamma_2,\phi)$ move through space with a velocity $\omega = (\omega_1,\omega_2) \in \mathbb{R}^2$ for a duration $\delta > 0$. The volume of the intersection of two slated elliptical cylinders (see Figure 1(b)) is given by

$$V(s,t,s',t') = (\delta - |t-t'|)_+ \times \nu_2(E(s,\gamma_1,\gamma_2,\phi) \cap E(\tilde{s},,\gamma_1,\gamma_2,\phi)),$$

where $\tilde{s} = (\tilde{s_1},\tilde{s_2})$ with $\tilde{s_i} = s_i' - |t'-t| \times \omega_i$, $i = 1,2$, where $\nu_d(\cdot)$ is the Lebesgue measure on $\mathbb{R}^d$.

For two fixed locations, the strength of dependence in the random field $\Lambda(x)$ is an increasing monotone function of the intersection volume; other choices of A are possible, provided that we are able to calculate efficiently the volume of the intersection. To efficiently calculate the ellipse intersection area, we use an approach for finding the overlap area between two ellipses, which does not rely on proxy curves; see Hughes and Chraibi (2012).[1]

In the sequel, we consider the measure

$$\alpha(B) = \alpha\nu_d(B)/\nu_d(A), \quad B \in \mathcal{B}_b(\mathcal{X}). \tag{9}$$

It follows that $\Lambda(x) \sim \text{Gamma}(\alpha,\beta)$, as required for model (4). Exploiting the formulas of Appendix A, the univariate Laplace transform of $\Lambda(x)$ is

$$\mathrm{LP}_x^{(1)}(v) := \mathrm{E}\left(e^{-v\Lambda(x)}\right) = \left(\frac{\beta}{v+\beta}\right)^{\alpha}, \tag{10}$$

and the bivariate Laplace transform of $\Lambda(x)$ and $\Lambda(x')$ is

$$\begin{aligned}\mathrm{LP}_{x,x'}^{(2)}(v_1,v_2) &:= \mathrm{E}\left(e^{-v_1\Lambda(x)-v_2\Lambda(x')}\right)\\ &= \left(\frac{\beta}{v_1+\beta}\right)^{\alpha(A_x\setminus A_{x'})}\left(\frac{\beta}{v_1+v_2+\beta}\right)^{\alpha(A_x\cap A_{x'})}\\ &\quad\times\left(\frac{\beta}{v_2+\beta}\right)^{\alpha(A_{x'}\setminus A_x)}.\end{aligned} \tag{11}$$

[1]The code is open source and can be downloaded from *http://github.com/chraibi/EEOver*.

This model for $\Lambda(x)$ is stationary, but nonstationarity in Gamma marginal distributions and/or dependence can be generated by using nonstationary indicator sets A_x whose size and shape depends on x. More general sets A_x with finite Lebesgue volume $\nu_3(A_x)$ could be used for constructing $\Lambda(x) = \Gamma(A_x)$. In all cases, the intersecting volume $\nu_3(A_{x_1} \cap A_{x_2})$ tends to zero if $\|x_2 - x_1\| \to \infty$, which establishes the property of α-mixing over space and time for the processes $\Lambda(x)$ and $Y(x)$. This property is paramount to ensure consistency and asymptotic normality in the pairwise likelihood estimation that we consider in the following (see Huser and Davison 2014).

3. Joint Tail Behavior of the Hierarchical Process

Extremal dependence in a bivariate random vector (Z_1, Z_2) can be explored based on the tail behavior of the conditional distribution $\Pr(Z_1 > F_1^{\leftarrow}(q) | Z_2 > F_2^{\leftarrow}(q))$ as q tends to 1, where $F_i^{\leftarrow}$, $i = 1, 2$ denotes the generalized inverse distribution functions of Z_i (Sibuya 1960; Coles, Heffernan, and Tawn 1999). The random vector (Z_1, Z_2) is said to be asymptotically dependent if a positive limit χ, referred to as the tail correlation coefficient, arises

$$\chi(q) := \frac{\Pr(Z_1 > F_1^{\leftarrow}(q), Z_2 > F_2^{\leftarrow}(q))}{\Pr(Z_2 > F_2^{\leftarrow}(q))} \to \chi > 0, \quad q \to 1^-.$$

The case $\chi = 0$ characterizes asymptotic independence.

To obtain a finer characterization of the joint tail decay rate under asymptotic independence, faster than the marginal tail decay rate, Coles, Heffernan, and Tawn (1999) introduced the $\overline{\chi}$ index defined through the limit relation

$$\bar{\chi}(q) := \frac{2 \log \Pr(Z_2 > F_2^{\leftarrow}(q))}{\log \Pr(Z_1 > F_1^{\leftarrow}(q), Z_2 > F_2^{\leftarrow}(q))} - 1 \to \bar{\chi} \in (-1, 1], \qquad q \to 1^-.$$

Larger values of $|\bar{\chi}|$ correspond to stronger dependence. We now show that $\{Z(x), x \in \mathcal{X}\}$ is an asymptotic independent process, that is, for all pairs $(x, x') \in \mathcal{X}^2$ with $x \neq x'$ the bivariate random vectors $(Z(x), Z(x'))$ are asymptotically independent.

Owing to the stationarity of the process, it is easy to show that for any $(x, x') \in \mathcal{X}^2$, $x \neq x'$ and for values v exceeding a threshold $u \geq 0$, we get

$$\begin{aligned}\Pr(Z(x) > v) &= \mathrm{LP}_x^{(1)}(v - u + \kappa) \\ &= \left(1 + \frac{v - u + \kappa}{\beta}\right)^{-\alpha(A_x)}\end{aligned}$$

and

$$\begin{aligned}\Pr(Z(x) > v, Z(x') > v) &= \mathrm{LP}_{x,x'}^{(2)}(v - u + \kappa, v - u + \kappa) \\ &= \left(1 + \frac{v - u + \kappa}{\beta}\right)^{-\alpha(A_x \setminus A_{x'})} \\ &\quad \times \left(1 + \frac{2v - 2u + 2\kappa}{\beta}\right)^{-\alpha(A_x \cap A_{x'})} \\ &\quad \times \left(1 + \frac{v - u + \kappa}{\beta}\right)^{-\alpha(A_{x'} \setminus A_x)}.\end{aligned}$$

To simplify notations, we set $c_0 := \alpha(A_x)$, $c_1 := \alpha(A_x \setminus A_{x'})$, $c_2 := \alpha(A_x \cap A_{x'})$, $c_3 := \alpha(A_{x'} \setminus A_x)$, such that $c_1 = c_3 = c_0 - c_2 \geq 0$ and $c_1 + 2c_2 + c_3 = 2c_0$. For $c_2 = 0$ characterizing disjoint indicator sets A_x and $A_{x'}$, it is clear that $Z(x)$ and $Z(x')$ are independent. Now, assume $u = 0$ without loss of generality and $x \neq x'$; then,

$$\begin{aligned}\chi_{x,x'}(v) &:= \frac{\Pr(Z(x) > v, Z(x') > v)}{\Pr(Z(x') > v)} \\ &= \left(1 + \frac{2v + 2\kappa}{\beta}\right)^{-c_2} \left(1 + \frac{v + \kappa}{\beta}\right)^{-c_1 - c_3 + c_0} \\ &= \left(1 + \frac{2v + 2\kappa}{\beta}\right)^{-c_2} \left(1 + \frac{v + \kappa}{\beta}\right)^{2c_2 - c_0} \\ &\sim 2^{-c_2} \left(\frac{v}{\beta}\right)^{c_2 - c_0}, \qquad \text{for large } v.\end{aligned}$$

Since $c_2 < c_0$, we obtain

$$\chi_{x,x'} = 0.$$

We conclude that Z is an asymptotic independent process.

To characterize the faster joint tail decay, we calculate

$$\begin{aligned}&\bar{\chi}_{xx'}(v) \\ &:= \frac{2 \log \Pr(Z(x) > v)}{\log \Pr(Z(x) > v, Z(x') > v)} - 1 \\ &= \frac{-2c_0 \log(1 + (v + \kappa)/\beta)}{-c_1 \log(1 + (v + k)/\beta) - c_2 \log(1 + 2(v + k)/\beta) - c_3 \log(1 + (v + k)/\beta)} - 1 \\ &= \frac{2c_0}{c_1 + c_2 \dfrac{\log(1 + 2(v + k)/\beta)}{\log(1 + (v + k)/\beta)} + c_3} - 1.\end{aligned}$$

Taking the limit for $v \to \infty$ yields

$$\bar{\chi}_{x,x'} = \frac{2c_0}{c_1 + c_2 + c_3} - 1 = \frac{c_2}{2c_0 - c_2},$$

which describes the ratio between the intersecting volume of A_x and $A_{x'}$ and the volume of the union of these two sets. The value of $\bar{\chi}$ confirms the asymptotic independence of the process Z. A larger intersecting volume between A_x and $A_{x'}$ corresponds to stronger dependence.

4. Composite Likelihood Inference

To infer the tail behavior of the observed data process $\{Z(x)\}$, without loss of generality assumed to have GP marginal distributions with shape parameter α, we focus on values exceeding a fixed high threshold u. We let θ denote the vector of unknown parameters. For simplicity, we assume that we have observed the excess values $Y(s_i, t)$ for a factorial design of S locations s_i, $i = 1, \ldots, S$ and T times $t = 1, \ldots, T$.

To exploit the tractability of intersecting volumes of two kernel sets, we focus on pairwise likelihood for efficient inference in our high-dimensional space-time set-up. The

pairwise (weighted) log-likelihood adds up the contributions $f(Y(s_i,t),Y(s_j,t+k);\theta)$ of the censored observations $Y(s_i,t),Y(s_j,t+k)$ and can be written

$$\mathrm{pl}(\theta)=\sum_{t=1}^{T}\mathrm{pl}_t(\theta)=\sum_{t=1}^{T}\sum_{k=0}^{\Delta_T}\sum_{i=1}^{S}\sum_{j=1}^{S}\{1-1_{\{i\geq j,k=0\}}\}\times\log f(Y(s_i,t),Y(s_j,t+k);\theta)w_{s_i,s_j},\tag{12}$$

where w_{s_i,s_j} is a weight defined on $[0,\infty)$ (Bevilacqua et al. 2012; Davis, Klüppelberg, and Steinkohl 2013b; Huser and Davison 2014). We opt for a cut-off weight with $w_{s_i,s_j}=1$ if $||s_i-s_j||\leq\Delta_S$ and 0 otherwise, which bypasses an explosion of the number of likelihood terms and shifts focus to relatively short-range distances where dependence matters most. This also avoids that the pairwise likelihood value (and therefore parameter estimation) is dominated by a large number of intermediate-range distances where dependence has already decayed to (almost) nil.

The contributions $f(Y(x),Y(x');\theta)$ are given by

$$f(y_1,y_2;\theta)=\begin{cases}\frac{\partial^2}{\partial v_1\partial v_2}\mathrm{LP}^{(2)}_{x,x'}(v_1,v_2)J(y_1)J(y_2) & y_1>0,y_2>0\\ \left(-\frac{\partial}{\partial v_1}\mathrm{LP}^{(1)}(v_1)+\frac{\partial}{\partial v_1}\mathrm{LP}^{(2)}_{x,x'}(v_1,v_2)\right)J(y_1) & y_1>0,y_2=0\\ \left(-\frac{\partial}{\partial v_2}\mathrm{LP}^{(1)}(v_2)+\frac{\partial}{\partial v_2}\mathrm{LP}^{(2)}_{x,x'}(v_1,v_2)\right)J(y_2) & y_1=0,y_2>0\\ 1-2\mathrm{LP}^{(1)}(v_1)+\mathrm{LP}^{(2)}_{x,x'}(v_1,v_2) & y_1=0,y_2=0\end{cases}$$

with $v_i=(\kappa+1)\left(1+\xi^*y_i/\sigma^*\right)^{1/\xi^*}-1$ and $J(y_i)=\frac{\kappa+1}{\sigma^*}\left(1+\frac{\xi^*y_i}{\sigma^*}\right)^{1/\xi^*-1}$, $i=1,2$. We provide analytical expressions for $\mathrm{LP}^{(1)}$ and $\mathrm{LP}^{(2)}_{x,x'}$ in Appendix B.

Since the space-time random field $\{\Lambda(x)\}$ is temporally α-mixing, the maximum pairwise likelihood estimator $\widehat{\theta}$ can be shown to be asymptotically normal for large T under mild additional regularity conditions; see Theorem 1 of Huser and Davison (2014). The asymptotic variance is given by the inverse of the Godambe information matrix $\mathcal{G}(\theta)=\mathcal{H}(\theta)[\mathcal{J}(\theta)]^{-1}\mathcal{H}(\theta)$. Therefore, standard error evaluation requires consistent estimation of the matrices $\mathcal{H}(\theta)=\mathrm{E}(-\nabla^2\mathrm{pl}(\theta))$ and $\mathcal{J}(\theta)=\mathrm{var}(\nabla\mathrm{pl}(\theta))$. We estimate $\mathcal{H}(\theta)$ with $\widehat{\mathcal{H}}=-\nabla^2\mathrm{pl}(\widehat{\theta})$ and $\mathcal{J}(\theta)$ through a subsampling technique (Carlstein 1986), implemented as follows. We define B overlapping blocks $D_b\subset\{1,\ldots,T\}$, $b=1,\ldots,B$, containing d_b observations; we write pl_{D_b} for the pairwise likelihood (12) evaluated over the block D_b. The estimate of $\mathcal{J}(\theta)$ is

$$\widehat{\mathcal{J}}=\frac{T}{B}\sum_{b=1}^{B}\frac{1}{d_b}\nabla\mathrm{pl}_{D_b}(\widehat{\theta})\nabla\mathrm{pl}_{D_b}(\widehat{\theta})'.$$

The estimates $\widehat{\mathcal{H}}$ and $\widehat{\mathcal{J}}$ allow us to calculate the composite likelihood information criterion (Varin and Vidoni 2005)

$$\mathrm{CLIC}=-\mathrm{pl}(\hat{\theta})+\mathrm{tr}\{\hat{\mathcal{H}}^{-1}\hat{\mathcal{J}}\}$$

with lower values of CLIC indicating a better fit. Similar to Davison and Gholamrezaee (2012), we improve the interpretability of CLIC values through rescaling CLIC*=c CLIC by a positive constant c chosen to give a pairwise log-likelihood value $\mathrm{pl}(\theta)$ comparable to the log-likelihood under independence.

5. Simulation Study

We assess the performance of the pairwise composite likelihood estimator through a small simulation study. For each replication, we consider $S=30$ randomly chosen sites on $[0,1]\times[0,1]$ observed at time points $t=1,\ldots,T=2000$. The realizations of the Gamma random field are simulated by adapting the algorithm of Wolpert and Ickstadt (1998b). In the simulations, we fix parameters $\xi=1,\sigma=10$ and an exceedance probability of $1-p=0.2$. We focus on estimating dependence parameters while treating the margins as known. For estimation, we fix the site-dependent threshold u to an empirical quantile of order greater than p. Here, we fix $p=0.9$ corresponding to $\kappa=9$.

Two scenarios with different model complexity are considered, involving different specifications of the cylinder (see Table 1). scenario A uses a circle-based cylinder without velocity, while scenario B comes with a slated ellipse-based cylinder, yielding nonnull velocity. Technically, the model in scenario A is over-parameterized since the rotation parameter ϕ cannot change the volume of the cylinder.

Model parameters are estimated on 100 data replications using the composite likelihood approach developed in Section 4. We have considered a larger number of replications for some parameter combinations, but in general the number of 100 replications is enough to satisfactorily illustrate the estimation efficiency. The evaluation of $pl(\theta)$ depends on the choice of Δ_S and Δ_T, where greater values increase the computational cost. Results in the literature indicate that using as much as computationally possible or all of the pairs will not necessarily lead to an improvement in estimation owing to potential issues with estimation variance (see, e.g., Huser and Davison 2014). We have considered different values of Δ_T and have identified $\Delta_T=15$ as a good compromise for the estimation quality. The parameter Δ_S has been set to 1 which is large enough with respect to the spatial domain limits. Main results are illustrated in the boxplots in Figures 2 and 3.

When the cylinder is circle-based, that is, $\gamma_1=\gamma_2$, and without velocity (scenario A), the orientation parameter ϕ can take any value. In the simulation experiment we estimate all parameters without constraints, such that the optimization algorithm gives also an estimate of ϕ. It is reassuring to see in the boxplots of Figure 2 that the other parameters are still well estimated.

Results are fairly good for the scenario B where the velocity is nonnull. The estimates of the velocity present slightly higher variability, and the estimation of ω_2 appears slightly biased. On the other hand, the duration δ and the lengths of the semi-axes of the ellipse (γ_1 and γ_2) are still well estimated. The angle ϕ is well defined in scenario B, but it is still estimated with relatively high variability. This may seem as disappointing at first glance, but it may be due to the only moderate difference in the length of

Table 1. Design of the two simulation scenarios.

	Parameters					
Scenario	γ_1	γ_2	ϕ	δ	ω_1	ω_2
A	0.2	0.2	–	10	0.00	0.00
B	0.2	0.3	$\pi/4$	5	0.05	0.10

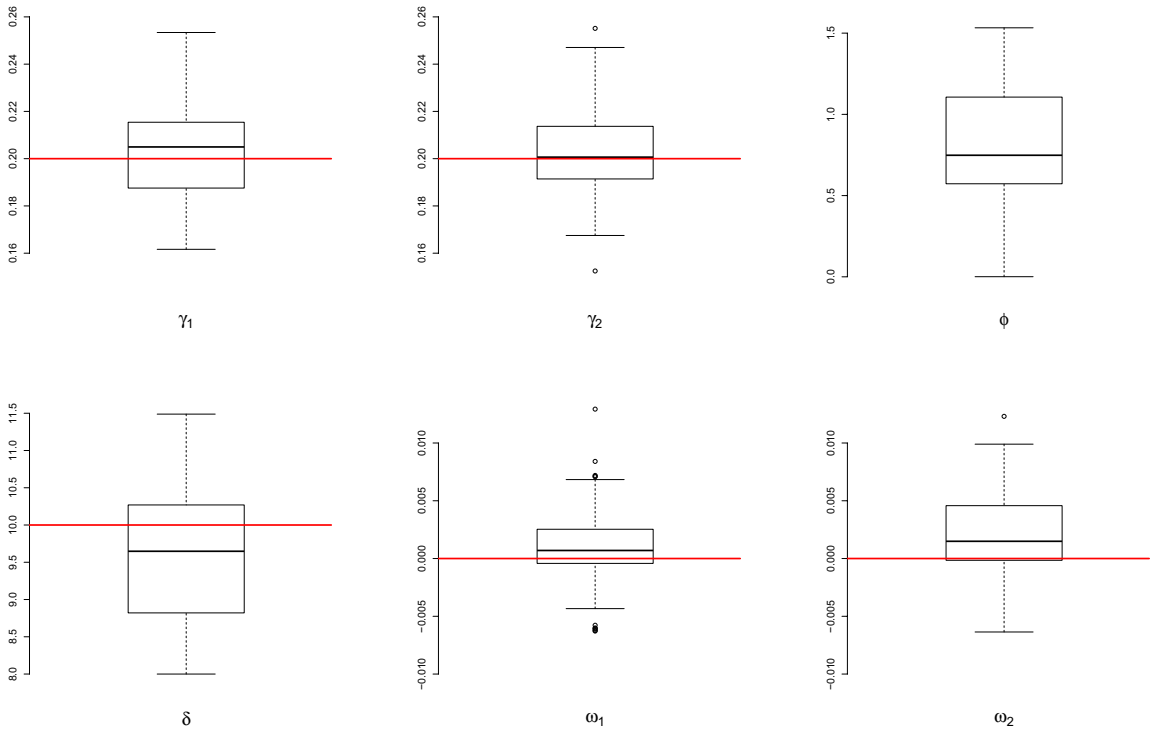


Figure 2. Summary of parameter estimates for scenario A of the simulation study: boxplots of parameter estimates for 100 simulated datasets.

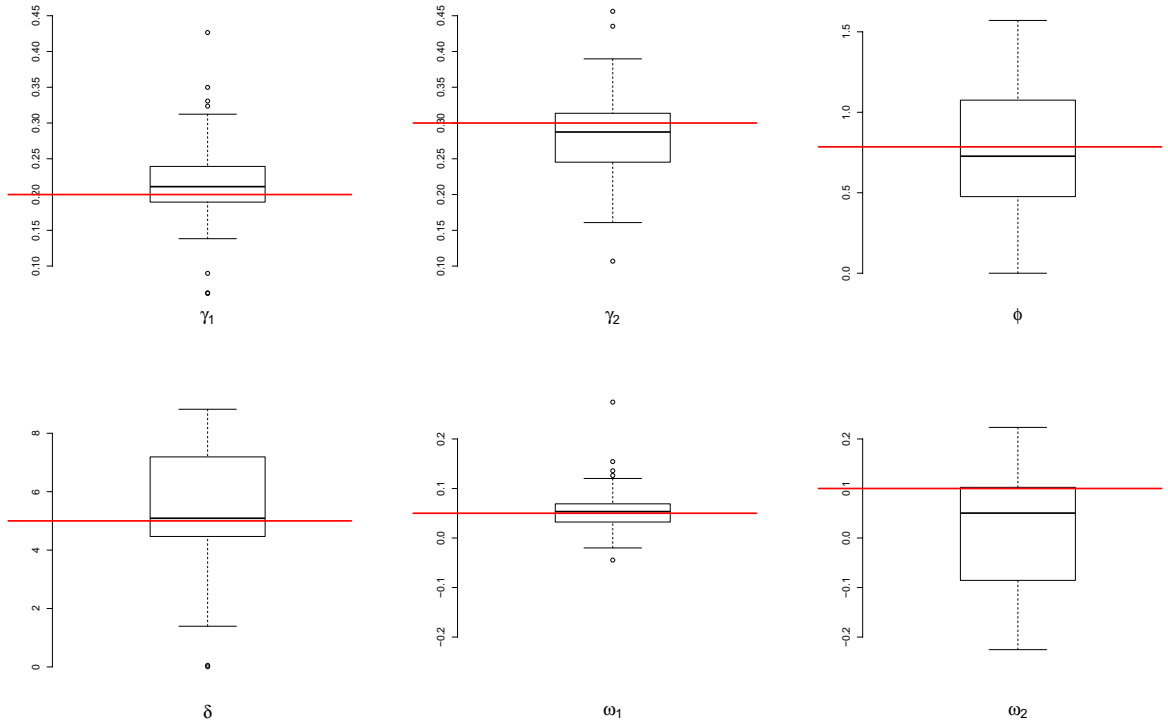


Figure 3. Summary of parameter estimates for scenario B of the simulation study: boxplots of parameter estimates for 100 simulated datasets.

the semi-axes. To check this conjecture, we consider a modified scenario B where the second semi-axis is modified from $\gamma_2 = 0.3$ to $\gamma_2 = 0.5$ and other parameters remain unchanged. As illustrated by the boxplots in Figure 4, estimation of ϕ clearly improves when the shape of the ellipse departs more strongly from a circular shape.

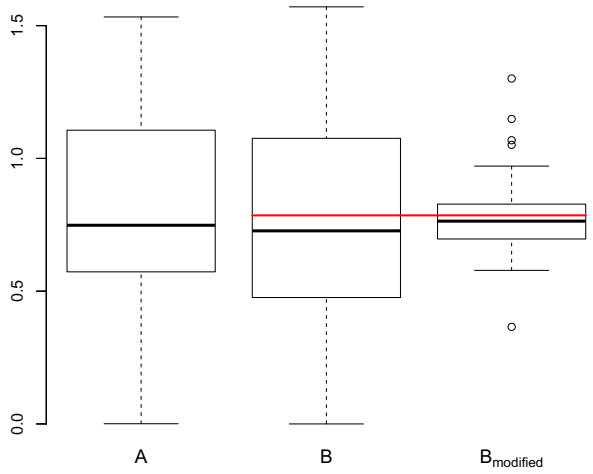


Figure 4. Parameter estimates of the simulation study: boxplots of ϕ estimates according to scenario A, scenario B, and a modified scenario B with $\gamma_2 = 0.5$.

Even with only a relatively small number of spatial sites and time steps, the simulation study shows that the pairwise composite likelihood approach leads to reliable estimates of model parameters that are well identifiable. We underline that results are consistently good whatever the complexity of the scenario.

6. Space-Time Modeling of Hourly Precipitation Data in Southern France

6.1. Data

We apply our hierarchical model to precipitation extremes observed over a study region in the South of France. Extreme rainfall events usually occur during fall season. They are mainly due to southern winds driving warm and moist air from the Mediterranean sea toward the relatively cold mountainous areas of the Cevennes and the Alps, leading to a situation which often provokes severe thunderstorms. The data were provided by Météo France (*https://publitheque.meteo.fr*). Our dataset is part of a query containing hourly observations at 213 rainfall stations for years 1993 to 2014. To avoid modeling complex seasonal trends, we keep only data from the September to November months, resulting in observations over 54,542 hr. For model fitting, we consider a subsample of 50 meteorological stations with elevations ranging from 2 to 1418 m, for which the observation series contain less than 70% of missing values over the full period. The spatial design of the stations is illustrated in Figure 5.

6.2. Exploratory Analysis

We fit the univariate model (5) for each station by fixing a threshold u that corresponds to the empirical 99% quantile. We use such a rather high probability value since we have many observations, and there is a substantial number of zero values such that a high quantile is needed to get into the tail region of the positive values. Figure 5 clearly shows that spatial nonstationarity arises in the marginal distributions.

Figure 6 displays the results of a bootstrap procedure in which we calculate estimates of $\chi(q)$ and $\bar{\chi}(q)$ for probabilities $q = 0.99, 0.995$ for pairs $Z(s,t), Z(s,t+h)$ with only temporal lag, and for pairs $Z(s,t), Z(s',t)$ with only spatial lag. The curves for spatial lags are the result of a smoothing procedure. Confidence bands are based on 200 bootstrap samples, drawn by the stationary bootstrap (Politis and Romano 1994). Our procedure samples temporal blocks of observations and the block length follows a geometric distribution with an average of 20 days. These plots support the assumption of asymptotic independence at all positive distances and at all positive temporal lags. Moreover, the strength of tail dependence as measured by the subasymptotic tail correlation value $\chi(q)$ strongly decreases when considering exceedances over increasingly high thresholds, which provides another clear sign of continuously decreasing and ultimately vanishing dependence strength. On the other hand, the values of the subasymptotic dependence measure $\overline{\chi}(q)$ (well adapted to asymptotic independence) decrease with increasing spatial distances or temporal lags, but they tend to stabilize at a nonzero value. This behavior indicates the presence of residual tail dependence that vanishes only asymptotically.

6.3. Modeling Spatio-Temporal Dependence

While the preceding exploratory analysis has shown that marginal distributions are not stationary, our model detailed in Section 2 requires a specific type of common marginal distributions. It would indeed be possible to extend the model to accommodate nonstationary patterns (an example can be found in Bortot and Gaetan (2016)) and to jointly estimate marginal and dependence parameters. However, our focus here is to illustrate that our modeling strategy is capable to capture complex stationary spatio-temporal dependence patterns at large values, which would render joint estimation of margins and dependence highly intricate. Therefore, we fit a GP distribution separately to each site with thresholds chosen as the empirical 99% quantile. With respect to positive precipitation, this quantile globally corresponds to a probability of 0.91, with a minimum of 0.86 and maximum of 0.95 over the 50 sites. Next, we use the estimated parameters $\hat{\xi}$ and $\hat{\sigma}$ to transform the raw exceedances $Y(x)$ observed at site x to exceedances $\tilde{Y}(x)$ with cdf (5) such that $\xi = 1$ and $\sigma = \kappa + 1$, that is,

$$\tilde{Y}(x) = (\kappa+1)\left\{\left(1+\frac{\hat{\xi}\,Y(x)}{\hat{\sigma}}\right)^{1/\hat{\xi}} - 1\right\}.$$

Since κ must satisfy $\Pr(\tilde{Y}(x) > 0) = (\kappa+1)^{-1} = 0.01$, see Equation (6), we get $\kappa = 99$.

We fit our hierarchical models to the censored pretransformed data $\tilde{Y}(x)$ by numerically maximizing the pairwise likelihood. We set the spatial cut-off distance to $\Delta_S = 110$ km, which retains about 60% of the pairs of meteorological stations, and we choose the temporal cut-off as $\Delta_T = 10$ hr. The resulting number of pairs of observations is approximately 4.6×10^9, taking into account missing values. The full pairwise likelihood counts around 1.7×10^{11} pairs, which shows that we have attained a huge reduction. Pairwise likelihood maximization is coded in C, and it runs in parallel using the

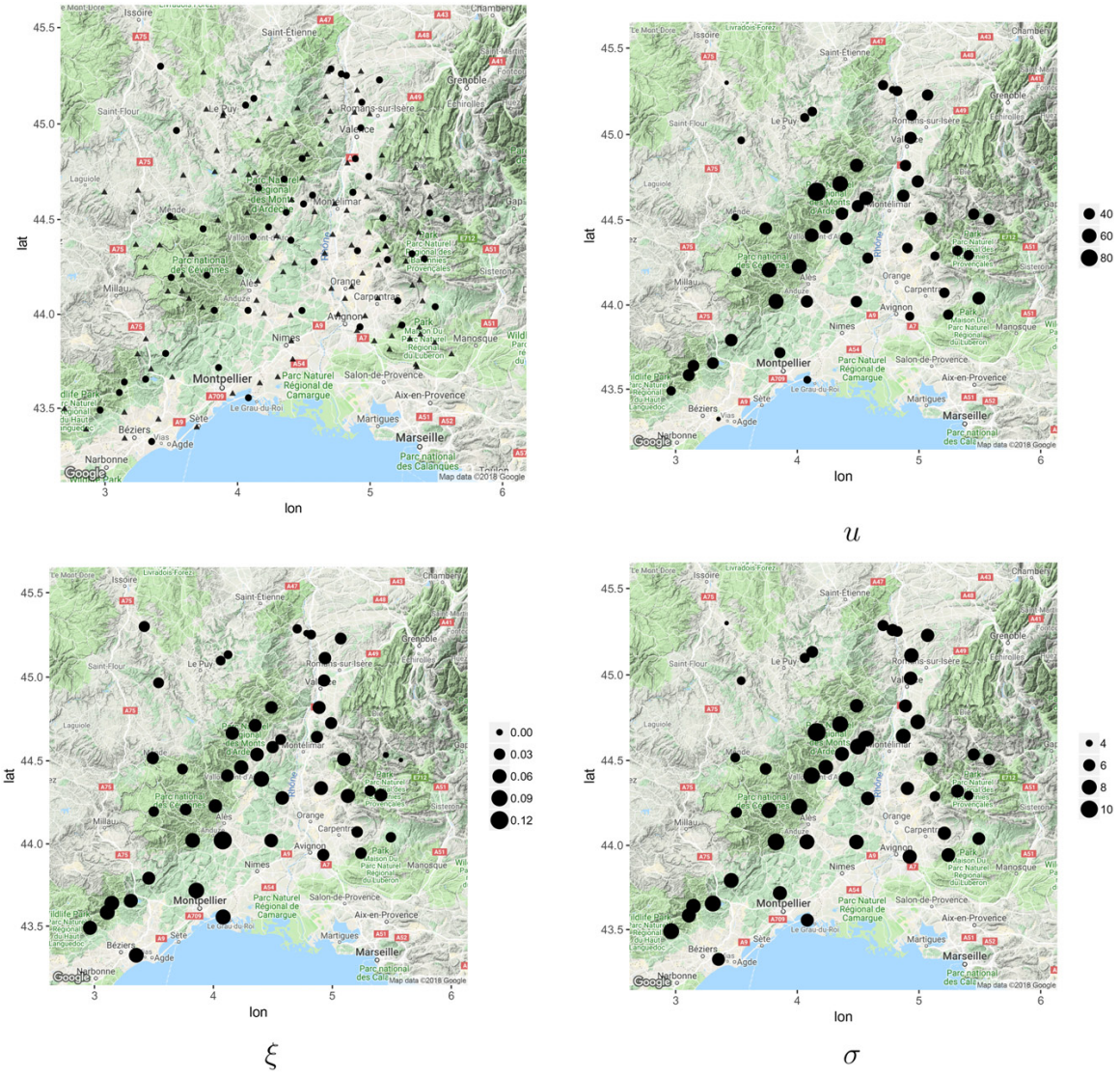


Figure 5. Precipitation data of Southern France. Top left display: topographic map showing the meteorological stations selected for our case study. Dots correspond to the stations used for fitting. In the other displays, their diameter is proportional to empirical 99% *quantiles* $u(s)$ (top right plot) and to estimates of the GPD parameters $\xi(s)$ (bottom left plot) and $\sigma(s)$ (bottom right plot).

R library `parallel`. All calculations were carried out on a 2.6 GHz machine with 32 cores and 52 GB of memory. One evaluation of the composite likelihood requires approximately 18 seconds. For calculating standard errors and CLIC* values, we use the previously described subsampling technique based on temporal windows by considering $B = 500$ overlapping blocks, each corresponding to 1000 consecutive hours, that is, $d_b = 50 \times 1000$.

We consider two settings for the hierarchical model, with (G1) and without velocity (G2). Then, we compare these two models to three variants of a censored Gaussian space-time copula model labeled C1, C2, and C3 (Bortot, Coles, and Tawn 2000; Renard and Lang 2007; Davison, Huser, and Thibaud 2013) pertaining to the class of asymptotic independent processes. The fits of the censored Gaussian space-time copula models match a censored Gaussian random field with transformed threshold exceedances; that is, we transform original data to standard Gaussian margins $G(x) = \Phi^{\leftarrow}(\mathrm{GP}(\tilde{Y}(x)))$ (with Φ the standard Gaussian cdf), and we suppose that $\{G(x), x \in \mathcal{X}\}$ is a Gaussian space-time random field with space-time correlation function $\rho(x_1, x_2; \theta)$.

We denote by $\rho_e(a) = \exp(-a)$ and by $\rho_s(a) = (1 - 1.5a + 0.5a^3)1_{[0,1]}(a)$, $a \geq 0$, the exponential and spherical correlation models with scale 1, respectively. We introduce the scaled Mahalanobis distance between spatial locations s_1 and s_2, written

$$a(s_1, s_2; \tau) = \{(s_1 - s_2)'\Omega(\tau)^{-1}(s_1 - s_2)\}^{1/2},$$

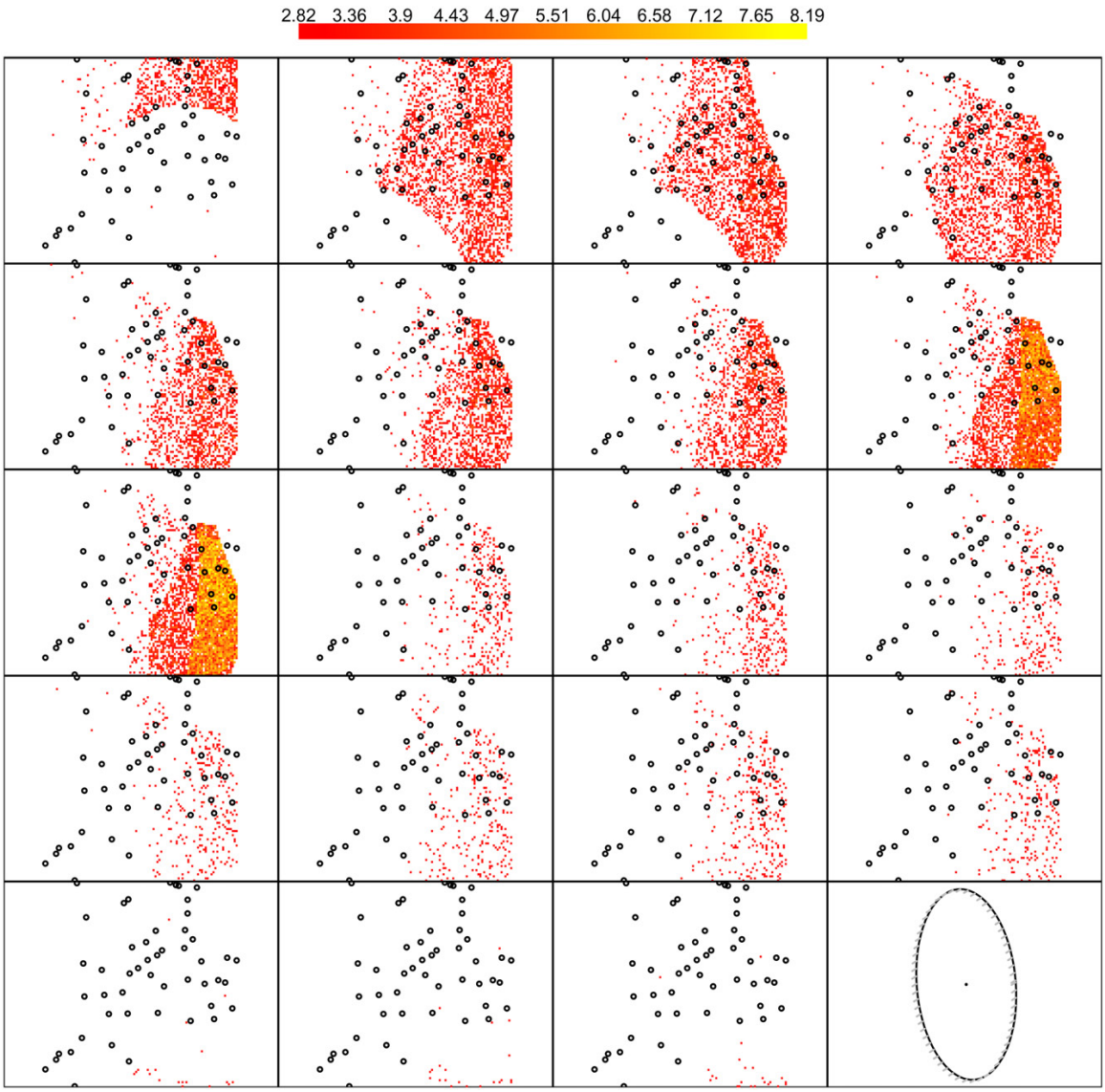


Figure 6. A simulation example showing exceedances of the 0.95-quantile for the model G1 fitted to precipitation data. Dots correspond to the stations used for fitting. The evolution over time during 19 hr is presented row-wise starting from the top left. The bottom right display illustrates the estimated ellipses, centered at the barycenter of the locations, and the movement induced by the velocity vector.

where

$$\Omega(\tau) = \begin{pmatrix} \cos(\tau_1) & -\sin(\tau_1) \\ \sin(\tau_1) & \cos(\tau_1) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \tau_2^{-1} \end{pmatrix} \times \begin{pmatrix} \cos(\tau_1) & \sin(\tau_1) \\ -\sin(\tau_1) & \cos(\tau_1) \end{pmatrix}.$$

The Mahalanobis distance defines elliptical isocontours. Here, $\tau_1 \in [0, \pi)$ is the angle with respect to the West-East direction, and $\tau_2 > 0$ is the length ratio of the two principal axes. We choose three specifications of the space-time correlation function:

C1. Space-time separable model:

$$\rho(x_1, x_2; \theta) = \rho_e(a(s_1, s_2; \tau)/\psi_S)\, \rho_e(|t_1 - t_2|/\psi_T), \quad (13)$$

with $\theta = (\tau_1, \tau_2, \psi_S, \psi_T)$. We assume anisotropic spatial correlation in analogy to models G1 and G2. The model is isotropic for $\tau_2 = 1$.

C2. Frozen field model 1 (see Christakos 2017, for a comprehensive account):

$$\rho(x_1, x_2; \theta) = \rho_e(a(s_1 - \nu t_1, s_2 - \nu t_2; \tau)/\psi), \quad (14)$$

where $\theta = (\tau_1, \tau_2, \psi, \nu')$ and $\nu \in \mathbb{R}^2$ is a velocity vector.

C3. Frozen field model 2 with compact support:

$$\rho(x_1, x_2; \theta) = \rho_s(a(s_1 - \nu t_1, s_2 - \nu t_2; \tau)/\psi). \quad (15)$$

In this model, two observations separated by Mahalonobis distance $a(s_1 - \nu t_1, s_2 - \nu t_2; \tau)$ greater than ψ will be independent.

Table 2. Estimates, standard errors (in italic), and CLIC* values of fitted models.

Model	Parameters						
	γ_1	γ_2	ϕ	δ	ω_1	ω_2	CLIC*
G1	165.062	318.823	0.085	20.184	0.723	0.446	404480.8
	23.459	*19.811*	*0.026*	*0.948*	*0.195*	*0.009*	
G2	175.817	294.323	0.041	20.036	0	0	404488.1
	11.879	*25.291*	*0.064*	*1.039*	–	–	
	τ_1	τ_2	ψ_S	ψ_T			CLIC*
C1	0.057	2.568	137.692	10.128			404626.2
	0.060	*0.309*	*7.615*	*0.523*			
	τ_1	τ_2	ψ_S		ν_1	ν_2	CLIC*
C2	1.034	2.025	108.755		6.672	16.358	404750.3
	0.040	*0.318*	*7.299*		*0.908*	*1.502*	
C3	0.481	5.125	174.980		6.614	10.406	405020.4
	0.005	*0.262*	*6.955*		*0.095*	*0.226*	

NOTES: Parameter units are kilometers for ϕ_S, γ_1, and γ_2, radians for ϕ and τ_1, hours for δ and ϕ_T, and kilometers per hour for ω_1, ω_2, ν_1, and ν_2.

Evaluation of the full likelihood of the models C1, C2, and C3 requires numerical operations such as matrix inversion, matrix determinants, and high-dimensional Gaussian cdfs (Genz and Bretz 2009), which are computationally intractable in our case. Therefore, we opt again for a pairwise likelihood approach, which also simplifies model selection through the CLIC*.

Estimation results are summarized in Table 2. The CLIC* in the last column shows a preference for our hierarchical models with the best value for model G1, followed closely by G2. Estimated durations vary only slightly between G1 and G2. Estimates of ϕ differ more strongly, but one has to take into account that estimates of both semi-axis are very close. Moreover, estimates of γ_1 and γ_2 are similar for G1 and G2, which suggests coherent results for the two models and allows reliable physical interpretation of estimated parameter values. Regarding the results for model G1, we observe that the estimated parameters γ_1 and γ_2 characterize an ellipse covering a large part of the study region, which indicates relatively strong dependence even between sites that are far separated in space.

The estimate of ϕ underlines the low inclination of the ellipse, while $\gamma_2 \approx 2\gamma_1$, which leads to an ellongated shape of the ellipse. It corresponds well to the orientation of the mountain ridges in the considered region.

The estimate of δ, which may be interpreted as the average duration of extreme events, corresponds well to empirical measures of the actual durations of extreme events in the study region. The orientation of the reliefs seems to play an important role for the estimated velocity characterized by the values of ω_1 and ω_2, with ω_1 being considerably larger than ω_2. For visual illustration, Figure 7 shows a simulation of model G1 where the velocity effect in precipitation intensities becomes apparent. This simulation shows heavy precipitation arriving from the north, predominantly spreading over the eastern slopes of a mountain range in the study region, and then becoming more intense and finally gradually evacuating toward the south.

Among the Gaussian copula models, the preference goes to the separable model C1.

To underpin the good fit of our models through visual diagnostics, Figure 8 shows estimated probabilities $\Pr(Z(s,t) > q | Z(s',t') > q)$ along different directions and at different temporal lags $|t - t'|$. These plots suggest that the behavior of models G1 and G2 is very close; there is no strong preference for one model over the other. The ranking of the copula models based on the CLIC* is also confirmed by the visual diagnostics. For contemporaneous observations with time lag 0, the models C1, C2, and C3 have comparable performance in capturing spatial dependence. However, for lags of 1 hr, models C2 and C3 represent the space-time interaction not satisfactorily.

Finally, we gain deeper insight into the joint tail structure of the fitted models by calculating empirical estimates $\hat{p}_i(h)$ of the multivariate conditional probability

$$\chi^*_{s_i;h}(q) := \Pr(Z(s_j,t) > q, s_j \in \partial s_i | Z(s_i, t-h) > q),$$

where ∂s_i is the set of the four nearest neighbors of site s_i, $i = 1, \ldots, 50$. We compare these values with precise Monte Carlo estimates $\tilde{p}_i^{(j)}(h)$, $j = 1, \ldots, 200$, based on a parametric bootstrap procedure using 200 simulations of the models G1, G2, and C1 with the leading CLIC* values. We compute site-specific root mean-squared errors (RMSE)

$$\mathrm{RMSE}_i(h) = \left\{ \frac{\sum_{j=1}^{200} (\tilde{p}_i^{(j)}(h) - \hat{p}_i(h))^2}{200} \right\}^{1/2},$$

as well as the resulting total RMSE, $\mathrm{RMSE}(h) = \sum_{i=1}^{50} \mathrm{RMSE}_i(h)$, as an overall measure of goodness of fit. Table 3 reports such values for fitted models using contemporaneous observations or lags of 1 or 2 hr ($h = 0, 1, 2$) between the reference site and its neighbors. If we consider the quantile $q_{0.99}$ used as a threshold for fitting models, our hierarchical models present the best fit in terms of RMSE only for lagged values. However, models G1 and G2 extrapolate better for larger values of the threshold such as $q_{0.995}$.

7. Conclusions

We have proposed a novel space-time model for threshold exceedances of data with asymptotically vanishing dependence strength. In the spirit of the hierarchical modeling paradigm with latent layers to capture complex dependence and time dynamics, it is based on a latent Gamma convolution process with nonseparable space-time indicator kernels, and therefore amenable to physical interpretation. This framework leads to marginal and joint distributions that are available in closed form and are easy to handle in the extreme value context. The assumption of conditional independence as in our model is practical since it avoids the need to calculate cumulative distribution functions in large dimensions, although difficulty remains in evaluating the volume of the intersections of more than two cylinders and in calculating partial derivates for full likelihood formula. We can draw an interesting parallel to the max-stable Reich–Shaby process $Z_{\mathrm{RS}}(x)$ (Reich and Shaby 2012), which is one of the more easily tractable spatial max-stable models and has a related construction. Indeed, the inverted process $1/Z_{\mathrm{RS}}(x)$ can be represented as the embedding of a dependent latent convolution process (based on positive α-stable variables) for the rate of an exponential distribution. Conditional independence models cannot accurately capture

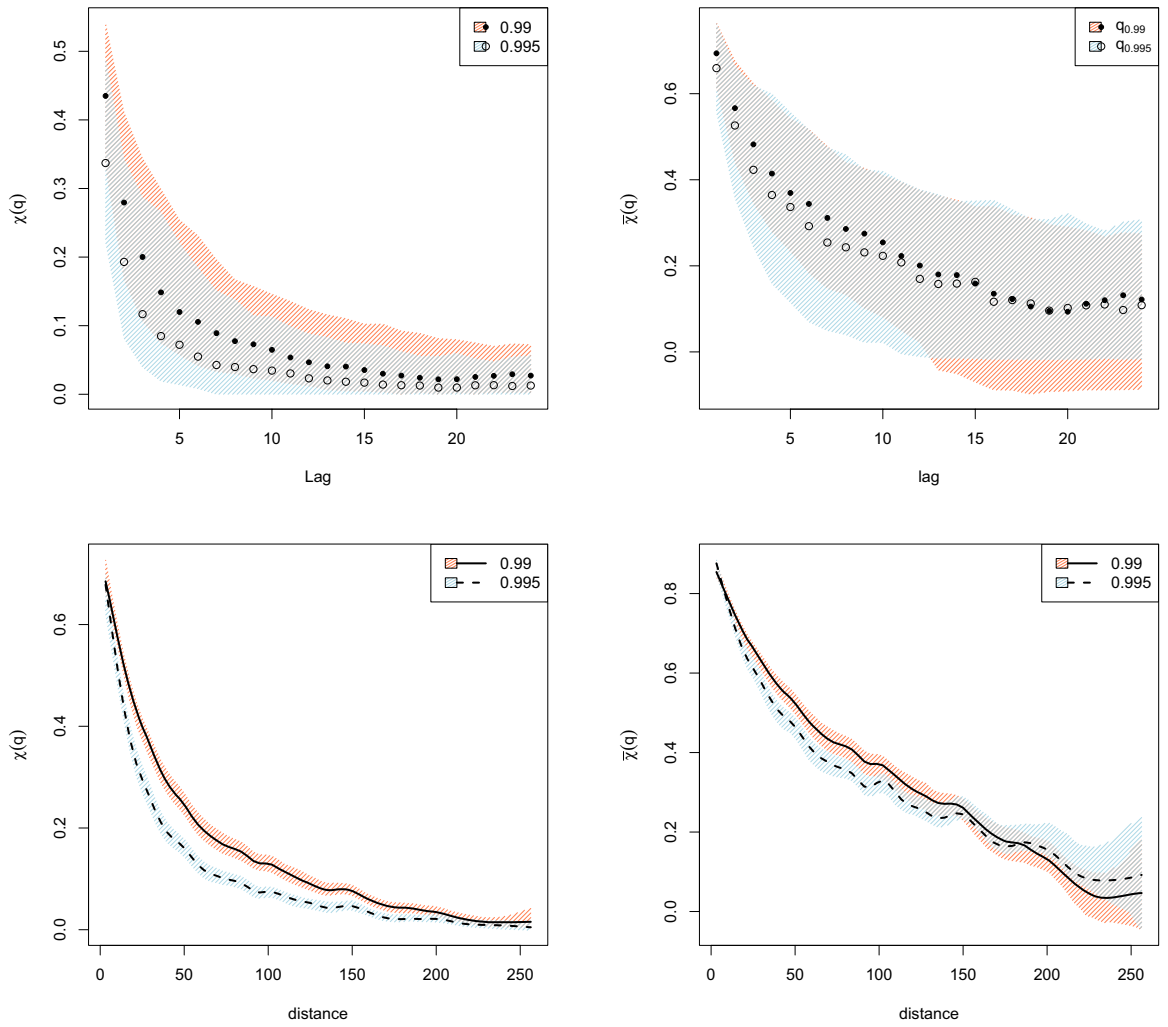


Figure 7. Empirical estimates of $\chi_x(q)$ (left panels) and $\bar{\chi}_x(q)$ (right panels) coefficients for the precipitation data. The filled region represents an approximate 95% confidence region based on a stationary bootstrap procedure.

the smoothness of the data-generating process. Nevertheless, the α-parameter in our model of the Gamma noise in (9) partially controls the smoothness of the latent Gamma field $\Lambda(s)$, with smaller values yielding more rugged surfaces.

In cases where data present asymptotic dependence, our asymptotically independent model may substantially underestimate the probability of jointly observing very high values over several space-time points. Asymptotic dependence in our construction (4) is equivalent to lower tail dependence in $\Lambda(x)$. There is no natural choice for introducing such dependence behavior, but a promising idea is to use what we label *Beta scaling*: given a temporal process $B(t)$ independent of $\Lambda(s,t)$ with Beta$(\tilde{\alpha},\alpha)$ distributed margins, $0 < \tilde{\alpha} < \alpha$, we could replace $\Lambda(s,t)$ in our construction by the process $\tilde{\Lambda}(s,t) = B(t)\Lambda(s,t)$ possessing margins following the $\Gamma(\tilde{\alpha},\beta)$ distribution. This construction has asymptotic dependence over space, and it will be asymptotically dependent over time if $B(t)$ has lower tail dependence. Follow-up work will explore theoretical properties and practical implementation of such extensions.

We have developed pairwise likelihood inference for our models, which scales well with high-dimensional datasets. We point out that handling observations over irregular time steps and missing data is straightforward with our model thanks to its definition over continuous time. While we think that MCMC-based Bayesian estimation of the relatively high number of parameters may be out of reach principally due to the very high dimension of the set of latent Gamma variables in the model's current formulation, we are confident that future efforts to tackle the conditional simulation of such space-time processes based on MCMC simulation with fixed parameters could be successful; that is, by using frequentist estimation of parameters, space-time prediction requires to iteratively update only the latent Gamma field through MCMC, but not parameters.

The application of our novel model to a high-dimensional real precipitation dataset from southern France was motivated

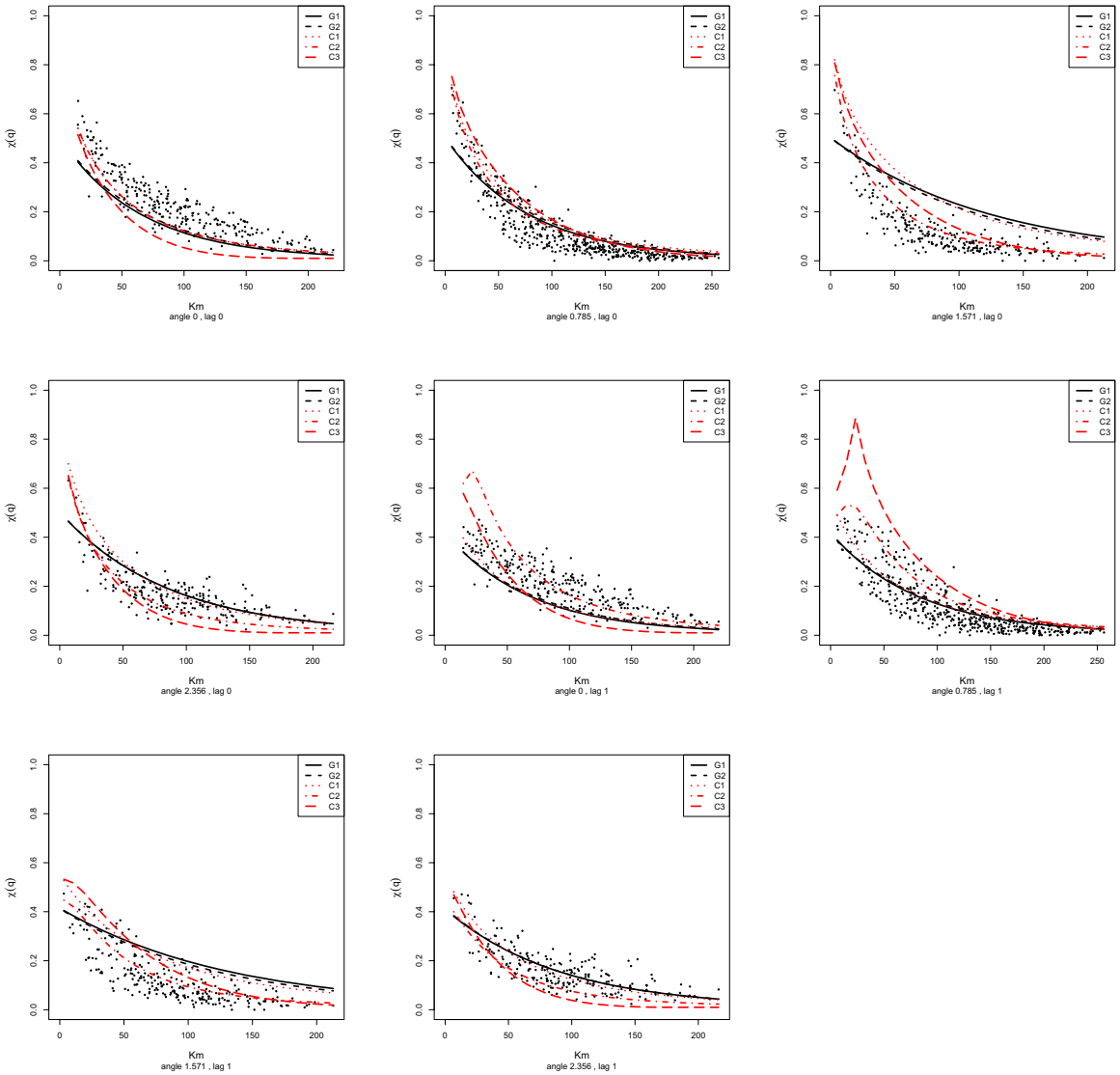


Figure 8. Estimated probabilities $\Pr(Z(s,t) > q | Z(s',t') > q)$ along different directions (expressed in radians) and at different temporal lags for the precipitation data. Dotted points correspond to empirical estimates. The value q is fixed to the empirical 99% quantile.

Table 3. Total root mean squared errors for the estimates of $\chi^*_{s_j;h}(q)$.

	RMSE(0)		RMSE(1)		RMSE(2)	
	$q_{0.99}$	$q_{0.995}$	$q_{0.99}$	$q_{0.995}$	$q_{0.99}$	$q_{0.995}$
G1	2.614	2.096	1.901	1.643	1.475	1.496
G2	2.605	2.072	1.907	1.626	1.477	1.480
C1	2.240	2.455	2.053	2.428	1.779	1.928

from clear evidence of asymptotic independence highlighted at an exploratory stage. It provides practical illustration of the high flexibility of our model and its capability to accurately predict extreme event probabilities for concomitant threshold exceedances in space and time. Based on meteorological knowledge about the precipitation processes in the study region, we had hoped to estimate a clear velocity effect. As a matter of fact, the fitted hierarchical model with velocity appeared to be only slightly superior to other models in some aspects. This interesting finding may also be interpreted as evidence for the highly fragmented structure arising in precipitation processes at small spatial and temporal scales.

Ongoing work aims to adapt the current latent process construction to the multivariate setting by considering constructions with Gamma factors common to several components, specifically structures with a hierarchical tree-based construction of the latent Gamma components, and extensions to asymptotic dependence using the above-mentioned Beta-scaling. Ultimately, such novelties could provide a flexible toolbox for multivariate space-time modeling with scenarios of partial or full asymptotic dependence.

Appendix A: Formulas for the Laplace Exponent of a Random Measure

The Laplace exponent of the random measure $\Gamma(\cdot)$ is defined as

$$\mathcal{L}(\phi) := -\log \mathrm{E}\left(\exp\left\{-\int \phi(x)\Gamma(dx)\right\}\right) = \int_{\mathcal{X}} \log\left\{1+\frac{\phi(x)}{\beta}\right\}\alpha(dx),$$

where ϕ is any positive measurable function.
Consider $\phi = \nu \mathbf{1}_A(x)$. Then,

$$\mathcal{L}(\phi) = -\log \mathrm{E}\left(\exp\{-\nu\Gamma(A)\}\right) = \int_A \log\left\{1+\frac{\nu}{\beta}\right\}\alpha(dx) = \alpha(A)\log\left\{1+\frac{\nu}{\beta}\right\},$$

that is,

$$\mathrm{E}\left(\exp\{-\nu\Gamma(A)\}\right) = \left(\frac{\beta}{\nu+\beta}\right)^{\alpha(A)}.$$

For bivariate analyses, choosing $\phi(x) = \nu_1 \mathbf{1}_{A_1}(x) + \nu_2 \mathbf{1}_{A_2}(x)$, yields

$$\begin{aligned}
\mathcal{L}(\phi) &= -\log \mathrm{E}\left(\exp\{-\nu_1\Gamma(A_1) - \nu_2\Gamma(A_2)\}\right)\\
&= -\log \mathrm{E}\left(\exp\{-\nu_1\Gamma(A_1\backslash A_2) - (\nu_1+\nu_2)\Gamma(A_1\cap A_2)\right.\\
&\quad \left.-\nu_2\Gamma(A_2\backslash A_1)\}\right)\\
&= \int_{A_1\backslash A_2} \log\left\{1+\frac{\nu_1}{\beta}\right\}\alpha(dx)\\
&\quad + \int_{A_1\cap A_2} \log\left\{1+\frac{\nu_1+\nu_2}{\beta}\right\}\alpha(dx)\\
&\quad + \int_{A_2\backslash A} \log\left\{1+\frac{\nu_2}{\beta}\right\}\alpha(dx)\\
&= \alpha(A_1\backslash A_2)\log\left\{1+\frac{\nu_1}{\beta}\right\} + \alpha(A_1\cap A_2)\log\left\{1+\frac{\nu_1+\nu_2}{\beta}\right\}\\
&\quad +\alpha(A_2\backslash A_1)\log\left\{1+\frac{\nu_2}{\beta}\right\}
\end{aligned}$$

and therefore

$$\begin{aligned}
&\mathrm{E}(\exp\{-\nu_1\Gamma(A_1) - \nu_2\Gamma(A_2)\})\\
&\quad = \left(1+\frac{\nu_1}{\beta}\right)^{-\alpha(A_1\backslash A_2)}\left(1+\frac{\nu_1+\nu_2}{\beta}\right)^{-\alpha(A_1\cap A_2)}\\
&\qquad \times \left(1+\frac{\nu_2}{\beta}\right)^{-\alpha(A_2\backslash A_1)}.
\end{aligned}$$

Appendix B: Formulas for the Pairwise Censored Likelihood

Let $\mathrm{LP}^{(1)}(\nu)$ and $\mathrm{LP}^{(2)}_{x,x'}(\nu_1,\nu_2)$, $x \neq x'$ denote the univariate and bivariate Laplace transform of $\Lambda(A_x)$ that is,

$$\mathrm{LP}^{(1)}(\nu) := \mathrm{E}\left(e^{-\nu\Lambda(A_x)}\right) = \left(\frac{\beta}{\nu+\beta}\right)^{c_0},$$

and

$$\mathrm{LP}^{(2)}_{x,x'}(\nu_1,\nu_2) := \mathrm{E}\left(e^{-\nu_1\Lambda(A_x)-\nu_2\Lambda(A_{x'})}\right) = \left(\frac{\beta}{\nu_1+\beta}\right)^{c_1} \times \left(\frac{\beta}{\nu_1+\nu_2+\beta}\right)^{c_2}\left(\frac{\beta}{\nu_2+\beta}\right)^{c_3}$$

with $c_0 = \alpha(A_x)$ $c_1 = \alpha(A_x\backslash A_{x'})$, $c_2 = \alpha(A_x\cap A_{x'})$, $c_3 = \alpha(A_{x'}\backslash A_x)$.

We obtain

$$\begin{aligned}
&\frac{\partial}{\partial\nu}\mathrm{LP}^{(1)}(\nu)\\
&\quad = -c_0\beta^{c_0}(\nu+\beta)^{-c_0-1},\\
&\frac{\partial}{\partial\nu_1}\mathrm{LP}^{(2)}_{x,x'}(\nu_1,\nu_2)\\
&\quad = -\beta^{c_1+c_2+c_3}\left\{c_1(\nu_1+\beta)^{-c_1-1}(\nu_1+\nu_2+\beta)^{-c_2}(\nu_2+\beta)^{-c_3}\right.\\
&\qquad \left. + c_2(\nu_1+\beta)^{-c_1}(\nu_1+\nu_2+\beta)^{-c_2-1}(\nu_2+\beta)^{-c_3}\right\},\\
&\frac{\partial}{\partial\nu_2}\mathrm{LP}^{(2)}_{x,x'}(\nu_1,\nu_2)\\
&\quad = -\beta^{c_1+c_2+c_3}\left\{c_3(\nu_1+\beta)^{-c_1}(\nu_1+\nu_2+\beta)^{-c_2}(\nu_2+\beta)^{-c_3-1}\right.\\
&\qquad \left. + c_2(\nu_1+\beta)^{-c_1}(\nu_1+\nu_2+\beta)^{-c_2-1}(\nu_2+\beta)^{-c_3}\right\},\\
&\frac{\partial}{\partial\nu_1\partial\nu_2}\mathrm{LP}^{(2)}_{x,x'}(\nu_1,\nu_2)\\
&\quad = \beta^{c_1+c_2+c_3}\left\{c_1c_2(\nu_1+\beta)^{-c_1-1}(\nu_1+\nu_2+\beta)^{-c_2-1}\right.\\
&\qquad \times(\nu_2+\beta)^{-c_3} + c_1c_3(\nu_1+\beta)^{-c_1-1}(\nu_1+\nu_2+\beta)^{-c_2}\\
&\qquad \times(\nu_2+\beta)^{-c_3-1} + c_2(c_2+1)(\nu_1+\beta)^{-c_1}(\nu_1+\nu_2+\beta)^{-c_2-2}\\
&\qquad \times(\nu_2+\beta)^{-c_3} + c_2c_3(\nu_1+\beta)^{-c_1}(\nu_1+\nu_2+\beta)^{-c_2-1}\\
&\qquad \left.\times(\nu_2+\beta)^{-c_3-1}\right\}.
\end{aligned}$$

Supplementary Materials

The supplementary material contains the code to simulate and estimate the model of Section 2. This is a snapshot of the repository *https://github.com/cgaetan/Gamma-GPD*. Real data cannot be freely distributed, but the Readme.md files contains instructions for requesting them.

Acknowledgments

The authors express their gratitude toward two anonymous referees and the associate editor for many useful comments that have helped improving earlier versions of the article. The authors thank Julie Carreau (IRD Hydro-Sciences, Montpellier, France) for helping them in collecting the data from the Meteo France database.

Funding

The work of the authors was supported by the French National Programme LEFE/INSU and by the LabEx NUMEV. Thomas Opitz acknowledges financial support from Ca' Foscari University, Venice, Italy.

References

Bacro, J. N., Gaetan, C., and Toulemonde, G. (2016), "A Flexible Dependence Model for Spatial Extremes," *Journal of Statistical Planning and Inference*, 172, 36–52. [2]

Barndorff-Nielsen, O. E., Lunde, A., Shepard, N., and Veraat, A. E. D. (2014), "Integer-Valued Trawl Processes: A Class of Stationary Infinitively Divisible Processes," *Scandinavian Journal of Statistics*, 41, 693–724. [2]

Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E. (2012), "Estimating Space and Space-Time Covariance Functions: A Weighted Composite Likelihood Approach," *Journal of the American Statistical Association*, 107, 268–280. [6]

Bortot, P., Coles, S., and Tawn, J. (2000), "The Multivariate Gaussian Tail Model: An Application to Oceanographic Data," *Journal of the Royal Statistical Society*, Series C, 49, 31–49. [2,9]

Bortot, P., and Gaetan, C. (2014), "A Latent Process Model for Temporal Extremes," *Scandinavian Journal of Statistics*, 41, 606–621. [2,3]

——— (2016), "Latent Process Modelling of Threshold Exceedances in Hourly Rainfall Series," *Journal of Agricultural, Biological, and Environmental Statistics*, 21, 531–547. [3,8]

Carlstein, A. (1986), "The Use of Subseries Values for Estimating the Variance of a General Statistic From a Stationary Sequence," *The Annals of Statistics*, 14, 1171–1179. [6]

Carreau, J., and Bouvier, C. (2016), "Multivariate Density Model Comparison for Multi-site Flood-Risk Rainfall in the French Mediterranean Area," *Stochastic Environmental Research Risk Assessement*, 30, 1591–1612. [1]

Casson, E., and Coles, S. G. (1999), "Spatial Regression Models for Extremes," *Extremes*, 1, 449–468. [1]

Christakos, G. (2017), *Spatiotemporal Random Fields Theory and Applications*, Amsterdam: Elsevier. [10]

Coles, S., Heffernan, J., and Tawn, J. (1999), "Dependence Measures for Extreme Value Analyses," *Extremes*, 2, 339–365. [5]

Cooley, D., Nychka, D., and Naveau, P. (2007), "Bayesian Spatial Modeling of Extreme Precipitation Return Levels," *Journal of the American Statistical Association*, 102, 824–840. [1]

Cox, D. R., and Isham, V. (1988), "A Simple Spatial-Temporal Model of Rainfall," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 415, 317–328. [2]

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013a), "Max-Stable Processes for Modeling Extremes Observed in Space and Time," *Journal of the Korean Statistical Society*, 42, 399–414. [2]

——— (2013b), "Statistical Inference for Max-Stable Processes in Space and Time," *Journal of the Royal Statistical Society*, Series B, 75, 791–819. [2,6]

Davis, R. A., and Mikosch, T. (2008), "Extreme Value Theory for Space-Time Processes With Heavy-Tailed Distributions," *Stochastic Processes and Their Applications*, 118, 560–584. [2]

Davison, A. C., and Gholamrezaee, M. M. (2012), "Geostatistics of Extremes," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 468, 581–608. [1,2,6]

Davison, A. C., Huser, R., and Thibaud, E. (2013), "Geostatistics of Dependent and Asymptotically Independent Extremes," *Journal of Mathematical Geosciences*, 45, 511–529. [1,2,9]

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012), "Statistical Modelling of Spatial Extremes," *Statistical Science*, 27, 161–186. [1]

de Haan, L., and Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, New York: Springer. [3]

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230. [3]

Ferreira, A., and de Haan, L. (2014), "The Generalized Pareto Process; With a View Towards Application and Simulation," *Bernoulli*, 20, 1717–1737. [1]

Gaetan, C., and Grigoletto, M. (2007), "A Hierarchical Model for the Analysis of Spatial Rainfall Extremes," *Journal of Agricultural Biological and Environmental Statistics*, 12, 434–449. [1]

Genz, A., and Bretz, F. (2009), *Computation of Multivariate Normal and t Probabilities*, New York: Springer. [11]

Hughes, G. B., and Chraibi, M. (2012), "Calculating Ellipse Overlap Areas," *Computing and Visualization in Science*, 15, 291–301. [4]

Huser, R., and Davison, A. C. (2014), "Space-Time Modelling of Extreme Events," *Journal of the Royal Statistical Society*, Series B, 76, 439–461. [2,5,6]

Huser, R., Opitz, T., and Thibaud, E. (2017), "Bridging Asymptotic Independence and Dependence in Spatial Extremes Using Gaussian Scale Mixtures," *Spatial Statistics*, 21, 166–186. [2]

——— (2018), "Penultimate Modeling of Spatial Extremes: Statistical Inference for Max-Infinitely Divisible Processes," arXiv no. 1801.02946. [2]

Huser, R., and Wadsworth, J. L. (2019), "Modeling Spatial Processes With Unknown Extremal Dependence Class," *Journal of the American Statistical Association*, 114, 414–434. [2]

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, London: Chapman & Hall. [3]

Kabluchko, Z., Schlather, M., and de Haan, L. (2009), "Stationary Max-Stable Fields Associated to Negative Definite Functions," *The Annals of Probability*, 37, 2042–2065. [1]

Morris, S. A., Reich, B. J., Thibaud, E., and Cooley, D. (2017), "A Space-Time Skew-t Model for Threshold Exceedances," *Biometrics*, 73, 749–758. [2]

Nieto-Barajas, L. E., and Huerta, G. (2017), "Spatio-Temporal Pareto Modelling of Heavy-Tail Data," *Spatial Statistics*, 20, 92–109. [2]

Noven, R. C., Veraart, A. E., and Gandy, A. (2015), "A Latent Trawl Process Model for Extreme Values," arXiv no. 1511.08190. [2]

Opitz, T. (2013), "Extremal t Processes: Elliptical Domain of Attraction and a Spectral Representation," *Journal of Multivariate Analysis*, 122, 409–413. [1]

——— (2016), "Modeling Asymptotically Independent Spatial Extremes Based on Laplace Random Fields," *Spatial Statistics*, 16, 1–18. [2]

——— (2017), "Spatial Random Field Models Based on Lévy Indicator Convolutions," arXiv no. 1710.06826. [2]

Opitz, T., Bacro, J.-N., and Ribereau, P. (2015), "The Spectrogram: A Threshold-Based Inferential Tool for Extremes of Stochastic Processes," *Electronic Journal of Statistics*, 9, 842–868. [1]

Politis, D. N., and Romano, J. P. (1994), "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313. [8]

Reich, B. J., and Shaby, B. A. (2012), "A Hierarchical Max-Stable Spatial Model for Extreme Precipitation," *The Annals of Applied Statistics*, 6, 1430–1451. [1,11]

Reiss, R., and Thomas, M. (2007), *Statistical Analysis of Extreme Values* (3rd ed.), Basel: Birkhäuser. [3]

Renard, B., and Lang, M. (2007), "Use of a Gaussian Copula for Multivariate Extreme Value Analysis: Some Case Studies in Hydrology," *Advances in Water Resources*, 30, 897 – 912. [9]

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987), "Some Models for Rainfall Based on Stochastic Point Processes," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 410, 269–288. [2]

Sang, H., and Gelfand, A. (2009), "Hierarchical Modeling for Extreme Values Observed Over Space and Time," *Environmental and Ecological Statistics*, 16, 407–426. [1,2]

Schlather, M. (2002), "Models for Stationary Max-Stable Random Fields," *Extremes*, 5, 33–44. [1,2]

Sibuya, M. (1960), "Bivariate Extreme Statistics," *Annals of the Institute of Statistical Mathematics*, 11, 195–210. [5]

Smith, R. L. (1990), "Max-Stable Processes and Spatial Extremes," Preprint, University of Surrey. [1]

Tawn, J., Shooter, R., Towe, R., and Lamb, R. (2018), "Modelling Spatial Extreme Events With Environmental Applications," *Spatial Statistics*, 28, 39–58. [1]

Thibaud, E., Mutzner, R., and Davison, A. C. (2013), "Threshold Modeling of Extreme Spatial Rainfall," *Water Resources Research*, 49, 4633–4644. [1]

Thibaud, E., and Opitz, T. (2015), "Efficient Inference and Simulation for Elliptical Pareto Processes," *Biometrika*, 102, 855–870. [1]

Varin, C., and Vidoni, P. (2005), "A Note on Composite Likelihood Inference and Model Selection," *Biometrika*, 52, 519–528. [6]

Wadsworth, J., and Tawn, J. (2012), "Dependence Modelling for Spatial Extremes," *Biometrika*, 99, 253–272. [2]

Wolpert, R. L., and Ickstadt, K. (1998a), "Poisson/Gamma Random Fields for Spatial Statistics," *Biometrika*, 85, 251–267. [2,3,4]

——— (1998b), "Simulation of Lévy Random Fields," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. Dey, P. Müller, and D. Sinha, New York: Springer, pp. 227–242. [6]

Annexe D

G1 - Extra-parametrized extreme value copula : extension to a spatial framework. Spatial Statistics (2020).

ARTICLE IN PRESS

Spatial Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Extra-parametrized extreme value copula : Extension to a spatial framework

J. Carreau [a,*], G. Toulemonde [b]

[a] HydroSciences Montpellier, CNRS/IRD/UM, Université de Montpellier - Case 17, 163 rue Auguste Broussonet 34090 Montpellier, France

[b] Institut Montpelliérain Alexander Grothendieck, Université de Montpellier, CNRS, Inria, France

ARTICLE INFO

Article history:
Received 22 February 2019
Accepted 15 January 2020
Available online xxxx

Keywords:
Gumbel copula
Spatial extremes
Heavy precipitation
ABC
Non-stationarity

ABSTRACT

Hazard assessment at a regional scale may be performed thanks to a spatial model for maxima that can be obtained by combining the generalized extreme-value (GEV) distribution for the univariate marginal distributions with extreme-value copulas to describe their dependence structure, as justified by the theory of multivariate extreme values. A flexible class of extreme-value copulas, called XGumbel for short, combines two Gumbel copulas with extra-parameters weighting each dimension. In a multisite study, the XGumbel copula quickly becomes over-parametrized. In addition, interpolation to ungauged locations is not easily achieved. We develop an extension of the XGumbel copula to the spatial framework by defining the extra-parameters as a mapping shaped as a disk. The inference of the Spatialized XGumbel copula is performed thanks to an Approximate Bayesian Computation (ABC) scheme with summary statistics based on upper tail dependence coefficients. The GEV parameters are estimated with a spatial regression model built with a vector generalized linear model. We evaluate and compare this spatial model with the Brown–Resnick process on annual maxima of daily precipitation totals at 177 gauged stations in the French Mediterranean over a 57 year period. Our analyses show that the ABC scheme yields, except in one instance, interpretable parameters. In addition, the Spatialized XGumbel copula is able to reproduce reasonably well the non-stationarity present in our case study.

* Corresponding author.
E-mail address: Julie.Carreau@ird.fr (J. Carreau).

https://doi.org/10.1016/j.spasta.2020.100410
2211-6753/

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

1. Introduction

The French Mediterranean is exposed to intense rainfall events called Cevenol events. These regularly cause flooding leading to important material damages and fatalities (Delrieu et al., 2005; Braud et al., 2014). Hazard assessment is conventionally performed by determining at-site T year return levels — the rainfall intensity level that is expected to be exceeded on average once per T years at a given site, see for instance Carreau et al. (2017). However, planning for flood risk mitigation is generally made at a regional scale. Therefore, a quantity of interest might rather be the probability that, conditionally on the fact that rainfall intensity at a given site has reached a high level, high intensity levels are likely to be reached at nearby sites. To estimate such a probability, characterization of the dependence of intense rainfall events in space, that is knowledge on spatial patterns of extreme events, is required. To this end, a spatial model for maxima over blocks of observations may be used.

Extreme value theory developed a sound theoretical framework to model the distribution of maxima over sufficiently large blocks of observations (Coles, 2001). Their univariate marginal distributions can be approximated by the Generalized Extreme Value (GEV) distribution (Fisher and Tippett, 1928; Gnedenko, 1943; Gumbel, 1958). In the multivariate case, theoretically justified distributions for componentwise maxima are the so-called Multivariate Extreme Value (MEV) distributions. The extension to the spatial setting leads to max-stable processes whose finite dimensional margins are MEV (de Haan, 1984). MEV distributions are either asymptotically dependent which entails that the dependence level remains constant at extreme levels or strictly independent (no dependence whatever the level).

MEV distributions and max-stable processes, unlike the GEV, do not have a unique finite dimensional parametrization (Beirlant et al., 2004). MEV distributions can be constructed by associating GEV margins with MEV copulas. Some MEV copulas such as the Gumbel copula exist in high dimension but are limited in their ability to reproduce complex dependencies. Moreover, interpolation to ungauged locations is not straightforward. Several parametric models for max-stable processes have been proposed, see Davison et al. (2012) for a recent review. For small study regions, a single parametric model may be used, for instance see Thibaud et al. (2013). However, in order to account for differences in dependence structures resulting from non-stationarities, larger study regions may be split into smaller sub-regions (Blanchet and Davison, 2011; Blanchet and Creutin, 2017).

As the complete log-likelihood is often intractable in high dimension, let alone in the spatial framework, pairwise log-likelihood inference is a common practice, in particular for max-stable processes (Davison et al., 2012). Another possibility is Approximate Bayesian Computation (ABC) likelihood free inference that selects parameters such that the model reproduces statistics of interest (see Beaumont, 2010 for instance). By simulating from the model for candidate parameters drawn from a prior distribution, ABC schemes constitute the so-called *reference table* that contains the statistics of interest. The posterior distribution consists of the candidate parameters that yielded statistics sufficiently similar to the observations'. ABC schemes for max-stable processes rely on summary statistics containing information on the extremal dependence structure (Erhardt and Smith, 2012; Erhardt and Sisson, 2016; Lee et al., 2018).

In this work, we propose a spatial model for maxima that rely on the extension to the spatial framework of the class of extra-parametrized MEV copulas (Durante and Salvadori, 2010; Salvadori and De Michele, 2010). The extra-parameters characterize each dimension thereby introducing additional flexibility in the dependence structure. We focus on extra-parametrized Gumbel (*XGumbel*) copulas, see Section 2. In Section 3, we present our case study, annual maxima of daily precipitation at 177 gauged stations over a 57 year period in the French Mediterranean. Our proposed spatial model, described in Section 4, combines the extension to the spatial framework of the XGumbel copula with a spatial regression model for the GEV marginals. This way, MEV distributions are defined for any set of sites, whether gauged or ungauged. The spatial extension is achieved by defining the extra-parameters of the XGumbel copula as a mapping of geographical covariates. An ABC scheme is designed to perform the inference. Evaluation on our precipitation case study is carried out in Section 5. The spatialized XGumbel copula is compared with the Brown–Resnick process, a max-stable process commonly used to model environmental extremes (Brown and Resnick, 1977; Davison et al., 2012).

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

2. Extra-parametrized Gumbel copula

2.1. Multivariate definition

The multivariate XGumbel copula $C_\psi(\cdot)$, defined as

$$C_\psi(\boldsymbol{u}) = C_{\beta_A}(\boldsymbol{u}^{\boldsymbol{a}})C_{\beta_B}(\boldsymbol{u}^{\boldsymbol{1-a}}), \quad \beta_A, \beta_B \geq 1, \ \boldsymbol{a} = (a_1, \ldots, a_d) \in [0,1]^d, \tag{1}$$

is a distribution function on the unit hypercube $[0,1]^d$ with parameter vector $\psi = (\beta_A, \beta_B, \boldsymbol{a})$. The parameters $\beta_A, \beta_B \geq 1$ are inherited from the two Gumbel copulas, $C_{\beta_A}(\cdot)$ and $C_{\beta_B}(\cdot)$, whose general form is

$$C_\beta(\boldsymbol{u}) = \exp\left\{-\left[\sum_{i=1}^{d}(-\ln u_i)^\beta\right]^{1/\beta}\right\}, \quad \beta \geq 1. \tag{2}$$

Note that the case $\beta = 1$ corresponds to the independent copula. As they affect all d dimensions in the same fashion, the two parameters β_A and β_B can be thought of as global parameters. The extra-parameter vector $\boldsymbol{a} = (a_1, \ldots, a_d) \in [0,1]^d$ appears as componentwise exponents in Eq. (1). As each dimension is characterized separately, extra-parameters may be thought of, in contrast to β_A and β_B, as local parameters. As can be seen from Eq. (1), if the values of β_A and β_B are swapped, the same copula C_ψ is obtained by replacing $\boldsymbol{a}$ with $\boldsymbol{1} - \boldsymbol{a}$. To remove this identifiability issue, we fix $\beta_A \leq \beta_B$.

As it fulfills the max-stability property, i.e. $C(u_1^t, \ldots, u_d^t) = C^t(u_1, \ldots, u_d)\ \forall t > 0$, the Gumbel copula is a multivariate extreme value (MEV) copula. By the definition in Eq. (1), it follows that C_ψ is a MEV copula as well (Salvadori and De Michele, 2010). The multivariate XGumbel copula may be obtained by a constructive approach as follows (see e.g. Liebscher, 2008). Let $\boldsymbol{U} \sim C_{\beta_A}$ and $\boldsymbol{V} \sim C_{\beta_B}$, then $\max(\boldsymbol{U}^{1/\boldsymbol{a}}, \boldsymbol{V}^{1/(\boldsymbol{1-a})})$ is distributed according to Eq. (1).

2.2. Bivariate properties

A MEV copula can be defined with the Pickands function conventionally denoted by A (Pickands, 1981; Marcon et al., 2017). In the bivariate case, a copula C is a MEV copula if and only if there exists a convex function $A : [0,1] \mapsto [1/2, 1]$ such that

$$\mathbb{P}(U_1 \leq u_1, U_2 \leq u_2) = C(u_1, u_2) = \exp\left[\ln(u_1u_2)A\left(\frac{\ln(u_2)}{\ln(u_1u_2)}\right)\right], \tag{3}$$

with U_1 and U_2 two uniform random variables on the interval $[0,1]$ and $0 \leq u_1, u_2 \leq 1$. The following properties must be fulfilled: $\min((1-t), t) \leq A(t) \leq 1$, for all $t \in [0,1]$, $A(0) = A(1) = 1$, $-1 \leq A'(0) \leq 0$, $0 \leq A'(1) \leq 1$ and $A'' \geq 0$.

For the bivariate XGumbel copula, the Pickands function, illustrated in Fig. 1(a), is

$$A(t) = \underbrace{\left[a_1^{\beta_A}(1-t)^{\beta_A} + a_2^{\beta_A}t^{\beta_A}\right]^{1/\beta_A}}_{A_{\beta_A}(t)} + \underbrace{\left[(1-a_1)^{\beta_B}(1-t)^{\beta_B} + (1-a_2)^{\beta_B}t^{\beta_B}\right]^{1/\beta_B}}_{A_{\beta_B}(t)} - 1. \tag{4}$$

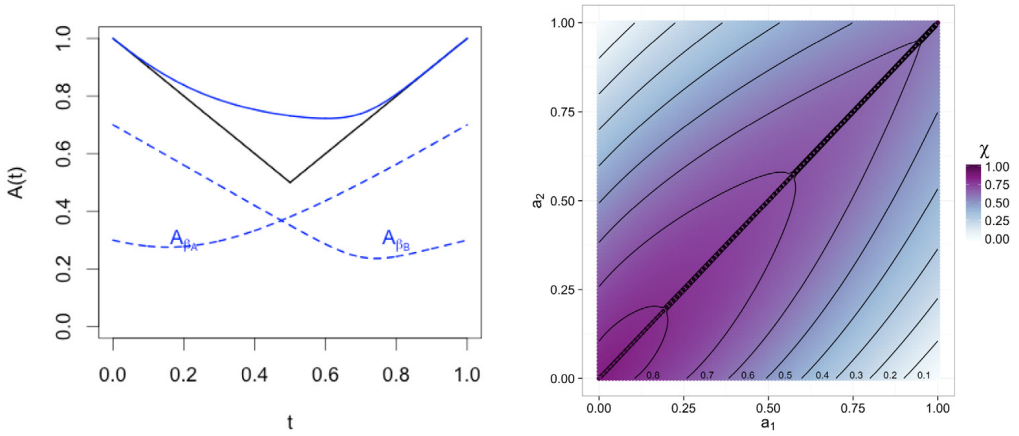
The Pickands function completely characterizes bivariate extremal dependence. It is equal to 1 in case of independence and equal to $\min((1-t), t)$ in case of perfect dependence. In between, the strength and the shape of the dependence, in particular the asymmetry, may vary. Note that the XGumbel copula is symmetrical when $a_1 = a_2$ or when $\beta_A = \beta_B$ and $a_1 = 1 - a_2$.

For 2-dimensional MEV distributions, the strength of extremal dependence may be summarized by the upper tail dependence coefficient χ defined as

$$\chi = \chi(u) = \mathbb{P}(U_2 > u \mid U_1 > u) = 2(1 - A(1/2)), \quad \forall\ 0 < u < 1. \tag{5}$$

In case of asymptotic independence, which necessarily corresponds to strict independence in a max-stable context, $\chi = 0$. Otherwise, $0 < \chi \leq 1$ indicates the strength of the asymptotic dependence (Sibuya, 1960; Coles et al., 1999).

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.



(a) *Pickands function of Eq. (4) with* $a_1 = 0.3$ *and* $a_2 = 0.7$.

(b) *Upper tail dependence coefficient* χ *of Eq. (6) with respect to* a_1 *and* a_2.

Fig. 1. Bivariate properties of the XGumbel copula of Eq. (1) with $\beta_A = 2$ and $\beta_B = 5$.

The upper tail dependence coefficient of the bivariate XGumbel copula is defined as

$$\chi = 2 - [(a_1^{\beta_A} + a_2^{\beta_A})^{1/\beta_A} + ((1-a_1)^{\beta_B} + (1-a_2)^{\beta_B})^{1/\beta_B}]. \tag{6}$$

It may be deduced by combining Eqs. (4) and (5). The variation of the χ of the XGumbel copula with respect to the values of the extra-parameters a_1 and a_2 is illustrated in Fig. 1(b) for $\beta_A = 2$ and $\beta_B = 5$. We note that χ is maximum when $a_1 = a_2$ (along the first diagonal) and increases for decreasing values of the extra-parameter (in the lower left corner). In the limiting case with $a_1 = a_2 = 0$ ($a_1 = a_2 = 1$), the XGumbel copula boils down to the Gumbel copula with parameter β_B (β_A) and $\chi = 2 - 2^{1/\beta_B}$ ($\chi = 2 - 2^{1/\beta_A}$). In addition, independence ($\chi = 0$) is achieved when $a_1 = 0$ and $a_2 = 1$ or the reverse, $a_1 = 1$ and $a_2 = 0$.

3. Precipitation data

3.1. Study area

Our study area is illustrated in Fig. 2. It covers about 16 000 km^2 around the city of Montpellier near the Mediterranean area in the south of France. It is well-known for intense rainfall events that occur mainly in autumn (Brunet et al., 2018). Owing to the Cévennes mountain range sitting in the north-west of the area, the Rhône river valley running in the eastern end that encompasses the city of Montpellier and the Mediterranean sea in the south, there is a strong variability in the distribution of heavy precipitation both in terms of intensities and of dependence structure (Blanchet and Creutin, 2017; Carreau et al., 2017).

We selected 177 gauged stations from the Météo-France network, the French weather service, that are located within our study area. For each station, we extracted annual maxima of daily precipitation totals over a 57 year period (1958–2014). The calibration set consists of the stations depicted as black filled circles. Among these, 11 numbered stations are used for a regional hazard analysis in Section 5.3. In addition, six stations with no missing values scattered in the study region, shown as red filled circles wearing letters in Fig. 2, are kept aside for validation purposes in Section 5.

3.2. Exploratory analyses of the dependence structure

We rely on sample estimates of the upper tail dependence coefficient χ introduced in Eq. (5) that summarizes the strength of the dependence between two sites i and j. Let U_i and U_j be random

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

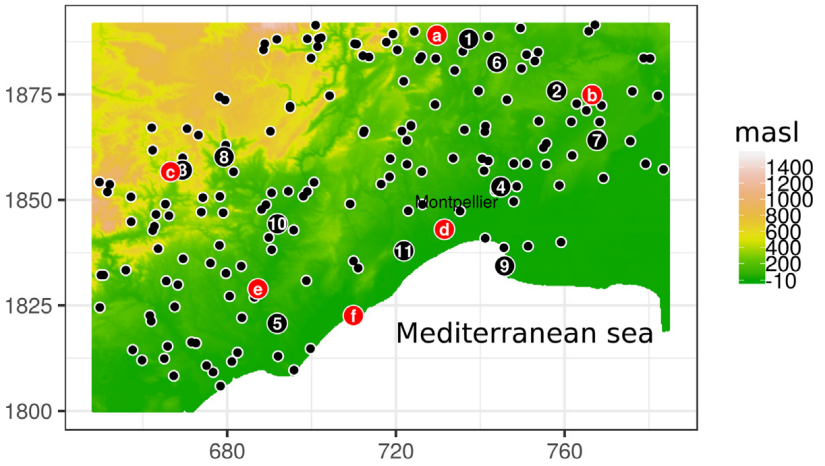


Fig. 2. Gauged stations in the study area located in the French Mediterranean: 171 stations (in black) are used for calibration, 11 of these are numbered and serve in a regional hazard analysis, 6 stations (in red) are kept aside for validation — coordinates are in extended Lambert II projection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

variables representing the annual maxima at each site transformed to the uniform scale. Then, the upper tail dependence coefficient χ_{ij} between sites i and j can be written as

$$\chi_{ij} = 2 - \left(\frac{1 + \mathbb{E}[|U_i - U_j|]}{1 - \mathbb{E}[|U_i - U_j|]} \right) \tag{7}$$

where $1/2\mathbb{E}[|U_i - U_j|]$ is the so-called madogram (Cooley et al., 2006; Vannitsem and Naveau, 2007). Sample estimates $\hat{\chi}_{ij}$ are obtained by replacing the expectation $\mathbb{E}[|U_i - U_j|]$ in Eq. (7) by the sample average. To compute empirical estimates, observed annual maxima are rank-transformed to the uniform scale by applying empirical distribution functions. For a given pair of stations, we kept empirical estimates only when at least 30 years of observations are available.

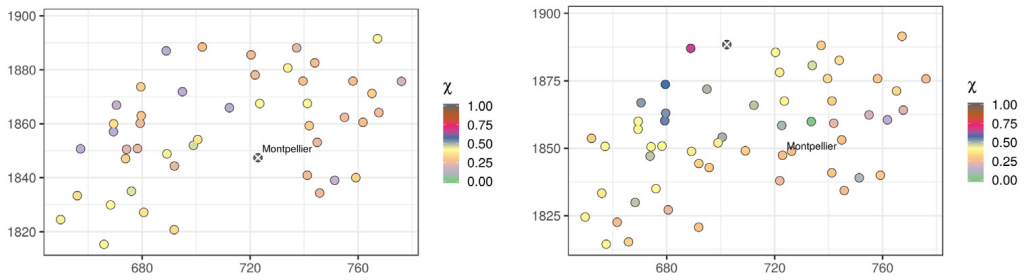
To assess the assumption of stationarity in the strength of the dependence, we depicted maps of estimates $\hat{\chi}_{ij}$, i being a fixed reference station and $j \in \{1, \ldots, 171\}$ being, in turn, each of the other calibration stations. In the left panel of Fig. 3, the reference station is the nearest one to the city of Montpellier which sits near the coastline. The strength of dependence is relatively low even for the closest stations ($\hat{\chi}_{ij}$ is about 0.3). In the right panel of Fig. 3, the reference station lies on the mountain range and the level of dependence is higher ($\hat{\chi}_{ij}$ is about 0.75 for the closest station). This change of dependence intensity with the location is an indication of non-stationarity.

We also assess the spatial behavior of the strength of dependence by looking at plots of estimates $\hat{\chi}_{ij}$ with respect to h, the distance between stations i and j. To reduce variability, we also computed estimates $\hat{\chi}_{[h]}$ for five classes of distance $[h] \in \{(0, 3], (3, 9], (9, 27], (27, 81], (81, 243]\}$ that follow a geometric progression. In Fig. 4, the pairwise estimates $\hat{\chi}_{ij}$ are shown (in gray) together with the distance class estimates $\hat{\chi}_{[h]}$ (in black) for the 171 stations of the calibration set over the 57 year period. Note that preliminary analyses performed by considering two orthogonal directions detected no significant anisotropy. The strength of dependence, as shown in Fig. 4, decreases with increasing distance, as is typical for extreme climatic spatial data (Blanchet and Davison, 2011; Davison and Gholamrezaee, 2012). However, the level of dependence seems to stabilize at a value clearly larger than zero, starting at a distance of about 40 km. This is an indication of asymptotic dependence.

4. Spatial XGumbel

The spatial XGumbel model for maxima presented in Section 4.3 combines a spatial regression model for the univariate marginal distributions introduced in Section 4.1 with the spatial extension

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.



(a) *The reference station i is near Montpellier.* **(b)** *The reference station i is on the mountain range.*

Fig. 3. Maps of empirical upper tail dependence coefficient estimates $\hat{\chi}_{ij}$ (see Eq. (7)) with respect to a given reference station i shown by a white cross. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

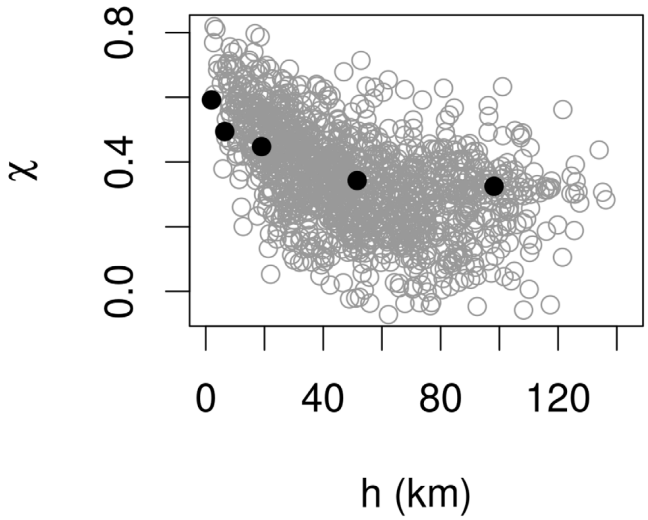


Fig. 4. Empirical upper tail dependence coefficient estimates $\hat{\chi}_{ij}$ for pairs of stations with at least 30 years of observations (in gray) and distance class estimates $\hat{\chi}_{[h]}$ (in black) with $[h] \in \{(0, 3], (3, 9], (9, 27], (27, 81], (81, 243]\}$ (see Eq. (7)).

of the multivariate XGumbel copula (see Section 2.1) in Section 4.2. A two-stage inference scheme for the spatial XGumbel model is described in Section 4.4.

4.1. Response surfaces

For a given site i, we denote Y_i as the random variable representing the annual maxima of daily precipitation. As is commonly done, we assume that G_i, the distribution function of Y_i, is the GEV distribution which has the form

$$G_i(y) = \exp\left[-\left\{1 + \xi_i\left(\frac{y - \mu_i}{\sigma_i}\right)\right\}_+^{-1/\xi_i}\right], \tag{8}$$

where $a_+ = \max(0, a)$. The GEV distribution, which is theoretically justified by the univariate extreme value theory (Fisher and Tippett, 1928; Gnedenko, 1943; Gumbel, 1958; Coles, 2001), depends on three parameters, see Eq. (8): the location parameter $\mu_i \in \mathbb{R}$, the scale parameter

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

$\sigma_i > 0$ and the shape parameter $\xi_i \in \mathbb{R}$. The latter characterizes the behavior of the upper tail of the distribution: exponential decay when $\xi_i = 0$, polynomial decay when $\xi_i > 0$ and finite endpoint for $\xi_i < 0$.

To obtain response surfaces that interpolate the GEV parameters over the study area, we rely on a vector generalized linear model (VGLM) approach, see Yee and Stephenson (2007) and Yee (2015). This allows to fit the GEV distribution simultaneously at all the calibration stations. The three GEV parameters are defined as functions

$$\mu(\boldsymbol{x}; \boldsymbol{\alpha_\mu}) = \alpha_{\mu:0} + \alpha_{\mu:1}x_1 + \cdots + \alpha_{\mu:p}x_p \tag{9}$$

$$\log(\sigma(\boldsymbol{x}; \boldsymbol{\alpha_\sigma})) = \alpha_{\sigma:0} + \alpha_{\sigma:1}x_1 + \cdots + \alpha_{\sigma:p}x_p \tag{10}$$

$$\log(\xi(\boldsymbol{x}; \boldsymbol{\alpha_\xi}) + 0.5) = \alpha_{\xi:0} + \alpha_{\xi:1}x_1 + \cdots + \alpha_{\xi:p}x_p, \tag{11}$$

where $\boldsymbol{x} \in \mathbb{R}^p$ are geographical covariates known everywhere in the study area. For the shape parameter, an offset of 0.5, see Eq. (11), serves to enforce that $\xi > -0.5$ thereby ensuring numerical stability (Yee, 2015).

4.2. Spatialized XGumbel copula

The spatialized XGumbel copula is based on the definition of the extra-parameters as a mapping $a : \mathbb{R}^2 \mapsto [0, 1]$, with parameters θ, of the x- and y-coordinates of the sites. Note that more general geographical covariates could be used as for the response surfaces. This mapping allows to extend the XGumbel copula from Eq. (1) to any set $\mathcal{S}$ of sites by letting the extra-parameters be given by $a_s = a(s_x, s_y; \theta)$, for all sites $s \in \mathcal{S}$ with x- and y-coordinates $(s_x, s_y) \in \mathbb{R}^2$. The vector of parameters ψ_{spat} of the spatialized XGumbel copula includes the global parameters β_A and β_B, as in Eq. (1), and θ to define the extra-parameter mapping. The number of parameters is thus invariant to the dimension, i.e. the number of sites in a spatial application. However, the extra-parameter mapping must be designed so that the resulting spatialized XGumbel copula be able to reproduce the spatial dependence structure of the observations.

To this end, we rely on the properties of the upper tail dependence coefficient χ of the XGumbel copula, see Fig. 1(b). First, we note that the dependence between two sites is maximized when their extra-parameter values are both equal to zero. In such a case, the extremal coefficient χ only depends on β_B (let $a_1 = a_2 = 0$ in Eq. (6)). Second, two sites are independent when one has extra-parameter value zero and the other has value one (let $a_1 = 0$ and $a_2 = 1$ or vice-versa in Eq. (6)). To account for these two points, we designed the extra-parameter mapping shaped as a disk, as shown in Fig. 5(a), with values approaching zero near the disk center indicating stronger dependence and values getting closer to one when moving away from the center implying independence between sites near the center and away from the center.

More precisely, for $(s_x, s_y) \in \mathbb{R}^2$, the extra-parameter mapping is parametrized as

$$a(s_x, s_y; \theta) = 1 - \exp\left\{-\frac{(s_x - \mu_x)^2 + (s_y - \mu_y)^2}{2\delta^2}\right\} \tag{12}$$

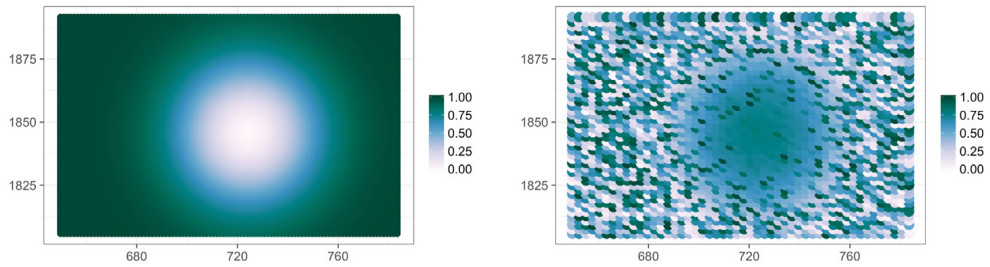
where $\delta > 0$ is a scale parameter and $(\mu_x, \mu_y) \in \mathbb{R}^2$ is the center of the disk. The extra-parameter mapping has thus three parameters $\theta = (\delta, \mu_x, \mu_y)$. Note that any pair of sites located in the dark green area in Fig. 5(a), whatever their distance, has the same dependence strength that only depends on β_A (let $a_1 = a_2 = 1$ in Eq. (6)). In order to permit independence between pairs of sites with larger distances, β_A is fixed to 1, i.e. the independent copula. Therefore, the parameter vector of the spatialized XGumbel copula is $\psi_{\text{spat}} = (\beta_B, \delta, \mu_x, \mu_y)$. In addition, the Pickands function and the upper tail dependence coefficient from Eqs. (4) and (6) are simplified as follows:

$$A(t) = a_1(1-t) + a_2 t + \left[(1-a_1)^{\beta_B}(1-t)^{\beta_B} + (1-a_2)^{\beta_B} t^{\beta_B}\right]^{1/\beta_B}$$

$$\chi = 2 - [(a_1 + a_2) + ((1-a_1)^{\beta_B} + (1-a_2)^{\beta_B})^{1/\beta_B}].$$

In Fig. 5(b), a simulation of the spatialized XGumbel copula reveals the impact on the spatial dependence pattern of the shape of extra-parameter mapping shown in Fig. 5(a). The area of strong

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.



(a) *Mapping with the disk centered at the city of Montpellier and scale parameter equal to 20 km.*

(b) *One simulation of the spatialized XGumbel copula with extra-parameter mapping from the left panel.*

Fig. 5. Effect of the shape of the extra-parameter mapping on the spatial pattern of a simulation of the spatialized XGumbel copula ($\beta_B = 20$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dependence is completely determined by the location of the disk center and the value of the scale parameter δ in Eq. (12). As areas of various degree of dependence may be defined, the spatialized XGumbel copula allows the introduction of non-stationarity in the dependence structure.

4.3. Proposed spatial model for maxima

The full spatial model for maxima combines the GEV distribution provided by the response surfaces in Section 4.1 and the spatialized XGumbel copula described in Section 4.2. For any set of sites, whether gauged or ungauged, this spatial model yields a well-defined MEV distribution. More precisely, ungauged sites can be modeled in a consistent way such that the lower dimensional distributions of sets of gauged and ungauged sites belong to the same class.

More precisely, let $\mathcal{S} = \{s_1, \ldots, s_K\}$ be any set of K sites in the study area, for any $K \in \mathbb{N}$. For all $s \in \mathcal{S}$ with x- and y-coordinates $(s_x, s_y) \in \mathbb{R}^2$, let Y_s and $\boldsymbol{x}_s$ be respectively the random variate representing the annual maxima of daily precipitation and the geographical covariates at site s. The GEV distribution function G_s, $\forall s \in \mathcal{S}$, has parameters $(\mu(\boldsymbol{x}_s; \boldsymbol{\alpha}_\mu), \sigma(\boldsymbol{x}_s; \boldsymbol{\alpha}_\sigma), \xi(\boldsymbol{x}_s; \boldsymbol{\alpha}_\xi))$ as provided by Eqs. (9)–(11). Moreover, the XGumbel copula parameter vector ψ contains the Gumbel copula parameters $\beta_A = 1$ and β_B that are shared for all the sites and the extra-parameters given by $a_s = a(s_x, s_y; \theta)$ from Eq. (12)$\forall s \in \mathcal{S}$. The multivariate distribution of the maxima at the K sites is then given by

$$\mathbb{P}(Y_{s_1} \leq y_1, \ldots, Y_{s_K} \leq y_K) = C_\psi(G_{s_1}(y_1), \ldots, G_{s_K}(y_K)), \tag{13}$$

with C_ψ defined in Eq. (1). Thanks to Eq. (13), it is possible to simulate from the model everywhere in the study area.

4.4. Inference scheme

As the joint estimation of the marginal and the dependence structure parameters of the spatial XGumbel model would be too complex, we opted for a two-stage inference scheme as follows. The parameter vectors $\boldsymbol{\alpha}_\mu$, $\boldsymbol{\alpha}_\sigma$ and $\boldsymbol{\alpha}_\xi$ of the response surfaces of the GEV parameters in Eqs. (9)–(11) are estimated by maximizing the log-likelihood under the independence assumption (Yee and Stephenson, 2007; Yee, 2015). The parameter vector ψ_{spat} of the spatialized XGumbel copula is estimated with an Approximate Bayesian Computation (ABC) scheme on the rank-transformed observations (as recommended in Genest and Favre, 2007).

To constitute the reference table of the ABC scheme, we use as summary statistics sample upper tail dependence coefficient estimates for distance classes $\hat{\chi}_{[h]}$ with $[h] \in \{(0, 3], (3, 9], (9, 27], (27, 81], (81, 243]\}$ based on the madogram (see Eq. (7)). In ABC schemes for max-stable processes,

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

related summary statistics containing information on the strength of the extremal dependence structure were proposed. In Erhardt and Smith (2012) and Erhardt and Sisson (2016), summary statistics deduced from the madogram and the extremal coefficient, which is equivalent to the upper tail dependence coefficient for max-stable distributions, were evaluated and compared. In addition to pairwise information, information based on triplet of sites was considered. A smoothing procedure, similar in spirits to the use of distance classes, was based on either curve fitting or by grouping stations.

The prior distribution in the ABC scheme of the spatialized XGumbel copula is meant to be vague. For the parameter vector $\psi_{\text{spat}} = (\beta_B, \delta, \mu_x, \mu_y)$, we set: $\beta_B \sim U[10, 100]$, $\delta \sim U[5100]$ and (μ_x, μ_y) is drawn uniformly from the locations of the 171 stations in the calibration set. The constitution of the reference table goes as follows, for all $i \in \{1, \ldots, 100000\}$:

1. Draw candidate parameters $\psi_{\text{spat}}^{(i)} = (\beta_B^{(i)}, \delta^{(i)}, \mu_x^{(i)}, \mu_y^{(i)})$ from the prior distribution ;
2. Simulate $\boldsymbol{U}^{(i)} = (U_1^{(i)}, \ldots, U_d^{(i)})$, a sample of size $n = 57$ from the spatialized XGumbel copula with parameters $\psi_{\text{spat}}^{(i)}$ at the $d = 171$ stations of the calibration set ;
3. Compute $\hat{\chi}_{[h]}$, the sample upper tail dependence coefficients for all $[h] \in \{(0, 3], (3, 9], (9, 27], (27, 81], (81, 243]\}$ on the simulated sample $\boldsymbol{U}^{(i)}$.

We apply a simple version of ABC called rejection-ABC in which the posterior distribution consists of a subset of candidate parameters such that the distance in terms of summary statistics to the observations is small. More specifically, let $\{\psi_{\text{spat}}^{(i_j)}\}_{j=1}^{100}$ with $1 \leq i_j \leq 100000$ be the subset of 100 candidate parameters such that Euclidean distances in terms of summary statistics are the smallest. This corresponds to 0.1% of the simulations from the prior distribution.

5. Assessment of spatial models for maxima

We evaluate and compare spatial models for maxima on the annual maxima of the daily precipitation data described in Section 3. In Section 5.1, a single spatial regression model for the univariate margins (see Section 4.1) is considered. In Section 5.2, the dependence structure as modeled by the spatialized XGumbel copula is compared with the one from a Brown–Resnick process (Brown and Resnick, 1977). The Brown–Resnick process is fitted by pairwise log-likelihood on the annual maxima rank-transformed to the Fréchet scale (this is performed with the `R` package from Ribatet, 2018). Uncertainty assessment is based on non-parametric bootstrap: 100 sets of Brown–Resnick parameters are estimated on bootstrap samples obtained by sampling with replacement the years of the calibration period. Finally, two complete spatial models, i.e. GEV margins combined with either the spatialized XGumbel copula or the Brown–Resnick process, are compared in Section 5.3 in terms of simulated fields of maxima and in terms of their ability to reproduce conditional trivariate probabilities involving the validation stations. These probabilities may be of interest for hazard assessment at a regional scale.

5.1. Response surfaces

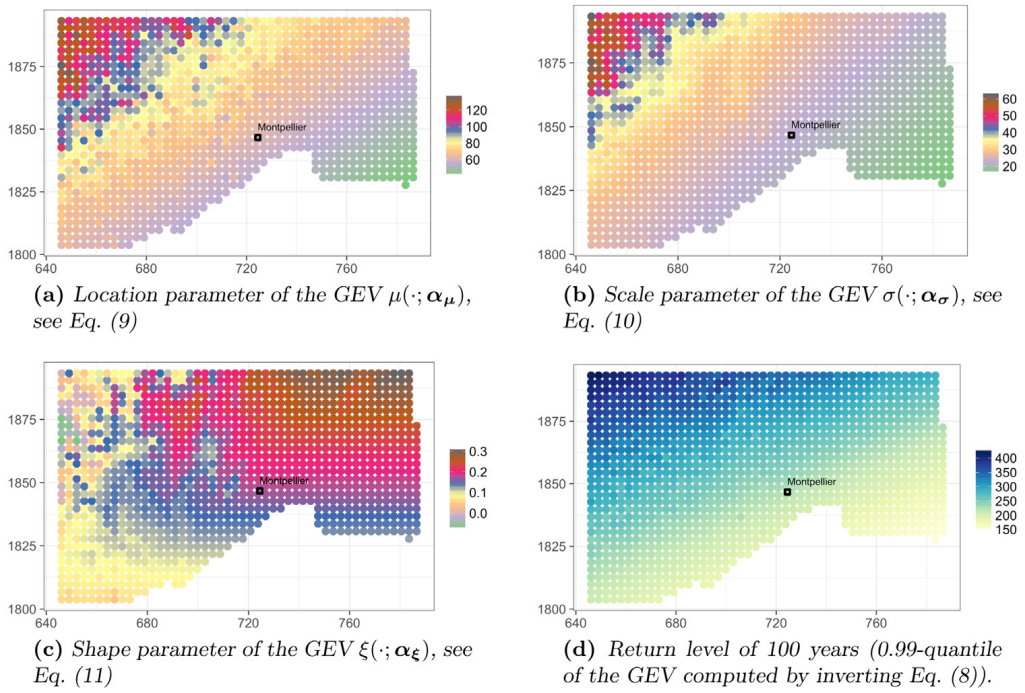
In addition to the x- and y-coordinates along with the altitude, we considered as covariates for the response surfaces in Eqs. (9)–(11) ten landscape features (Benichou and Le Breton, 1987). Based on a digital elevation model, these features are deduced from a principal component (PC) analysis applied to the relative elevation of a square neighborhood centered on each cell of the digital elevation grid. The first ten components are retained.

Covariate selection is performed in two stages. First, a screening is performed by applying LASSO regression to the natural logarithm of the annual maxima with the initial 13 covariates (Friedman et al., 2010). Six covariates are selected: the x- and y- coordinates, the altitude and three landscape features resulting from the 1st, 4th and 9th PC. This selection is further refined by constraining the coefficients of the VGLM to be null when not sufficiently significant for a subset of the GEV parameters. The Bayesian Information Criterion (BIC) is used to ensure that the exclusion of covariates does not deteriorate the fit (Schwarz, 1978). The final covariate selection is summarized in Table 1.

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

Table 1
Selected covariates for the response surfaces of the GEV parameters (Eqs. (9)–(11)). In addition to the x- and y-coordinates and the altitude z, three landscape features (PC1, PC4 and PC9) are obtained from a principal component analysis of the digital elevation grid.

	x	y	z	PC1	PC4	PC9
$\mu(\cdot;\boldsymbol{\alpha}_\mu)$	✓	✓	✓	✓	✓	✓
$\sigma(\cdot;\boldsymbol{\alpha}_\sigma)$	✓	✓	✓			
$\xi(\cdot;\boldsymbol{\alpha}_\xi)$		✓	✓			



(a) *Location parameter of the GEV* $\mu(\cdot;\boldsymbol{\alpha}_\mu)$*, see Eq. (9)*

(b) *Scale parameter of the GEV* $\sigma(\cdot;\boldsymbol{\alpha}_\sigma)$*, see Eq. (10)*

(c) *Shape parameter of the GEV* $\xi(\cdot;\boldsymbol{\alpha}_\xi)$*, see Eq. (11)*

(d) *Return level of 100 years (0.99-quantile of the GEV computed by inverting Eq. (8)).*

Fig. 6. Interpolation of the GEV parameters over a grid covering the study area with a vector generalized linear model approach, see Eqs. (9)–(11), and geographical covariates (see Table 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The response surfaces of the GEV parameters as provided by the fitted VGLM by interpolating over a grid covering our study area are shown in Figs. 6(a)–6(c). While the spatial patterns of the location and scale parameters are strongly influenced by the altitude, the shape parameter displays a different pattern with higher values in the Rhône river valley. The map of the 100-year return levels, i.e. quantiles of probability 0.99 computed by inverting Eq. (8), is shown in Fig. 6(d). As is typical for this area, values ranging from 150 mm near the coastline to 400 mm on the mountain range are observed (Carreau et al., 2017).

In Fig. 7, the goodness-of-fit of the response surfaces is evaluated in terms of return levels at the six validation stations. Each validation station, depicted in red filled circles in Fig. 2, wears a letter that is related to a panel in Fig. 7. Empirical return levels are depicted as black dots. The light blue bands are 99% non-parametric bootstrap confidence bands (10000 replications obtained by sampling with replacement the 57 years of annual maxima) for the return levels computed from the GEV parameters interpolated by the fitted VGLM. At the third station which is located in the mountain area (corresponding to the red filled circle wearing the letter c in Fig. 2), the VGLM interpolation tends to overestimate the larger empirical return levels (see Fig. 7(c)). Nevertheless, the fit is overall quite satisfactory.

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

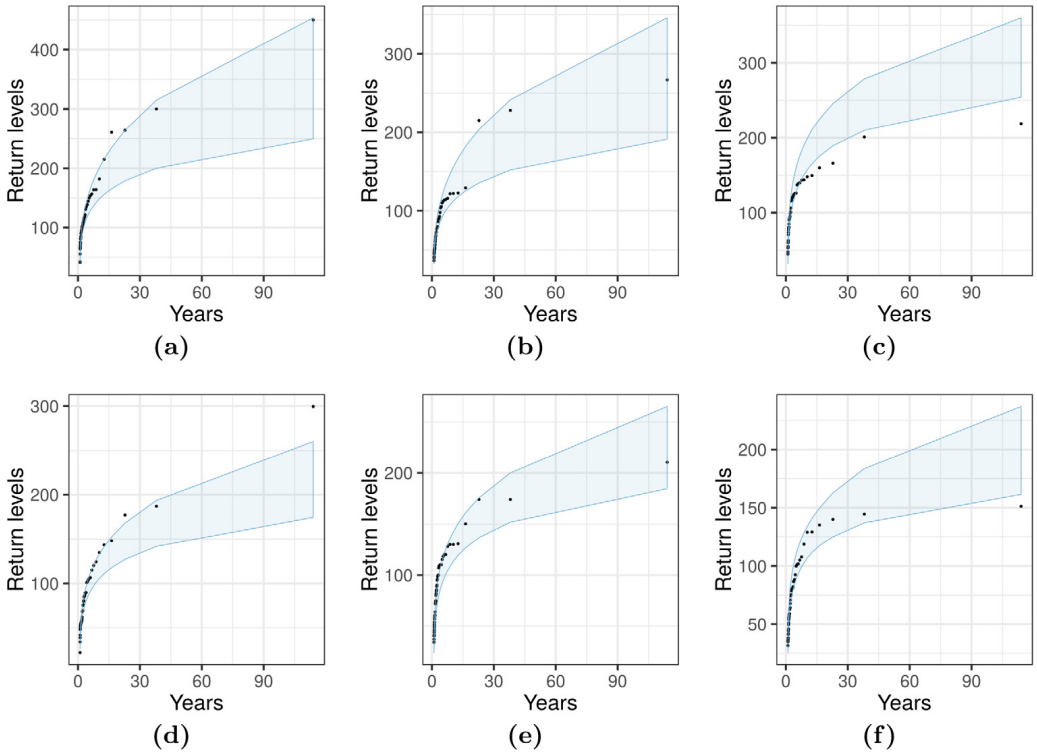


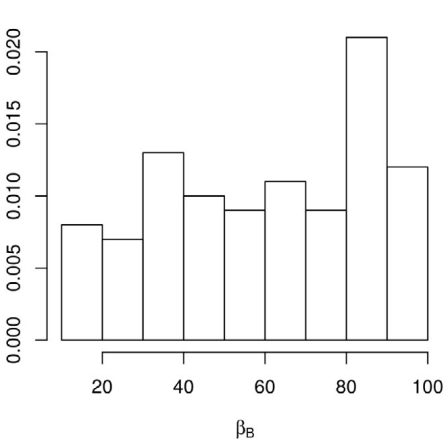
Fig. 7. Return levels at the six validation stations, each panel corresponding to a red filled circle wearing the same letter in Fig. 2: empirical estimates are depicted as black dots and 99% non-parametric bootstrap confidence bands for the return levels computed from GEV parameters interpolated by the fitted VGLM are shown in light blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2. Spatial dependence structures

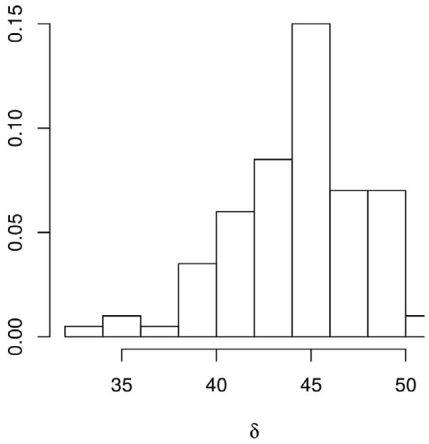
The posterior distribution of the spatialized XGumbel copula parameter vector resulting from the ABC scheme, that is the subset $\{\psi_{\text{spat}}^{(i_j)}\}_{j=1}^{100}$ with $1 \leq i_j \leq 100000$ of candidate parameters leading to the summary statistics closest to the observed ones, is illustrated in Fig. 8. For the Gumbel parameter β_B, in Fig. 8(a), the posterior distribution is similar to the prior distribution $U[10, 100]$. This might indicate that the designed ABC scheme is not able to infer properly this parameter. In contrast, the posterior distribution of δ, the scale parameter of the disk in the extra-parameter mapping, has a clear mode at about 45 km, see Fig. 8(b). The posterior distribution of the location of the disk center in the extra-parameter mapping is represented as black filled circles in Fig. 8(c). The selected disk centers are located preferentially, i.e. 98 times out of 100, over the mountain range, in a very specific area which might be explained by orographic effects.

The spatialized XGumbel copula and the Brown–Resnick process are compared in Fig. 9, left and right panel respectively, in terms of the statistics $\hat{\chi}_{[h]}$, i.e. the sample upper tail dependence coefficients for distance classes $[h]$, with $[h] \in \{(0, 3], (3, 9], (9, 27], (27, 81], (81, 243]\}$. The empirical estimates computed from the observations are shown in light blue in both panels. For each model, there are 100 statistics $\hat{\chi}_{[h]}$ depicted in gray. For the spatialized XGumbel copula, these statistics, retrieved directly from the reference table, correspond to the 100 sets of parameters $\{\psi_{\text{spat}}^{(i_j)}\}_{j=1}^{100}$ with $1 \leq i_j \leq 100000$ from the posterior distribution of the rejection-ABC inference scheme. The median of the 100 $\hat{\chi}_{[h]}$ is also shown in black. For the Brown–Resnick process, the 100 statistics are estimated by simulating samples of the same size as the observations' from

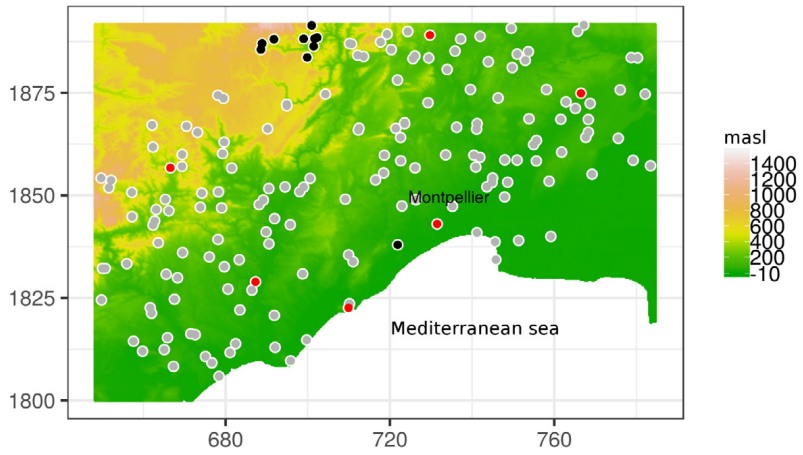
Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.



(a) *Gumbel parameter β_B (see Eq. (1)).*



(b) *Scale parameter δ of the disk in the extra-parameter mapping (see Eq. (12)).*



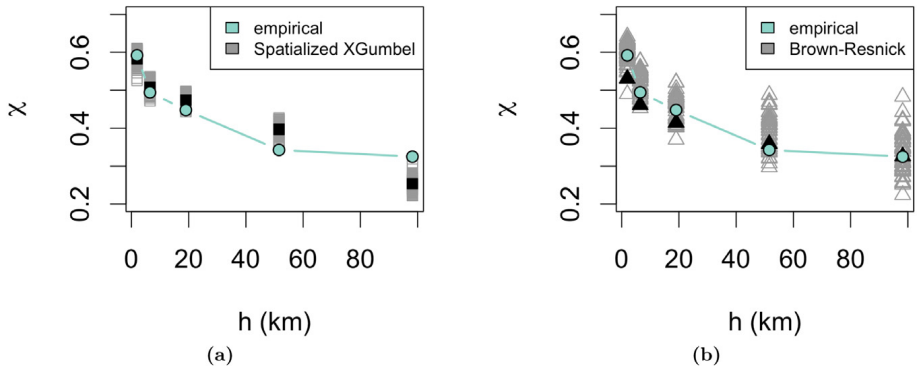
(c) *Location parameter (μ_x, μ_y) (black filled circles) of the center of the disk in the extra-parameter mapping (see Eq. (12)) : $^{98}/_{100}$ in the mountain area. Other calibration (validation) stations are shown in grey (red).*

Fig. 8. Posterior distribution of the spatialized XGumbel copula parameters $\{\psi_{\text{spat}}^{(i_j)}\}_{j=1}^{100}$ with $1 \leq i_j \leq 100000$ from the rejection-ABC inference scheme described in Section 4.4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the 100 sets of Brown–Resnick parameters obtained by non-parametric bootstrap. The statistics estimated from the fit on the original calibration data are shown in black. The patterns of decrease in extremal dependence with the distance produced by both models of spatial dependence structure are comparable to the one obtained from the observed annual maxima. However, the spread and thus the uncertainty of the Brown–Resnick estimates is larger.

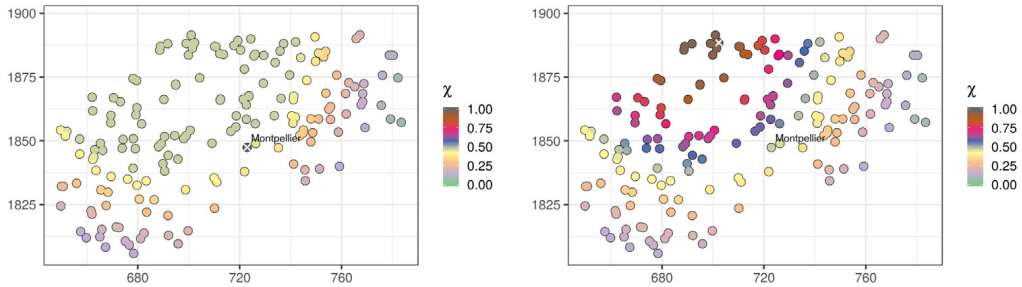
In Figs. 10 and 11, the two models are compared in terms of non-stationarity patterns in the dependence structure. These patterns are produced when drawing the maps of the upper tail

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.



(a) (b)

Fig. 9. Upper tail dependence coefficient estimates $\hat{\chi}_{[h]}$ for five classes of distance $[h] \in \{(0, 3], (3, 9], (9, 27], (27, 81], (81, 243]\}$ (see Eq. (7)). For the spatialized XGumbel (left panel), the best 100 estimates (in gray, with the median in black) are retrieved from the reference table. For the Brown–Resnick process (right panel), estimates are computed on samples of the same size as the observations' (57 years); from 100 models fitted on bootstrap samples (in gray) and from the model on the original sample (in black).. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) *The reference station i is near Montpellier.* **(b)** *The reference station i is on the mountain range.*

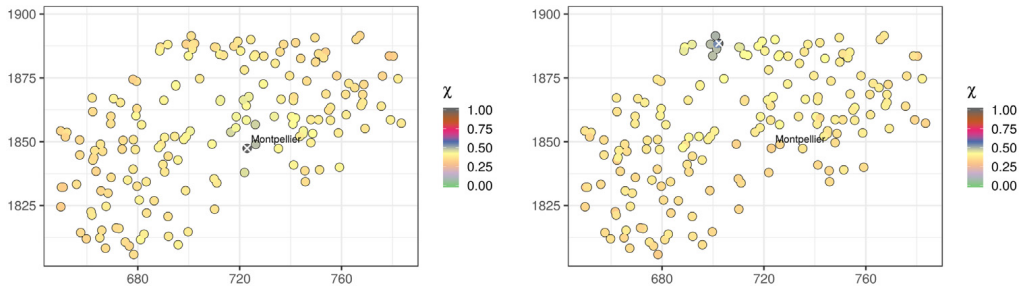
Fig. 10. Maps of spatialized XGumbel copula upper tail dependence coefficient estimates $\hat{\chi}_{ij}$, computed from Eq. (6). The reference station i is shown by a white cross. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dependence coefficient estimates $\hat{\chi}_{ij}$ with respect to two different reference sites i. In Fig. 10, these patterns are depicted for the spatialized XGumbel copula, with the $\hat{\chi}_{ij}$ obtained by replacing the parameters in Eq. (6) by the best set of parameters from the posterior distribution of the ABC scheme. For the Brown–Resnick process, the maps are shown in Fig. 11 with $\hat{\chi}_{ij}$ computed with the madogram, as in Eq. (7), on a sample of size 1000 simulated from the fitted model. Although the values are a bit too high with respect to the empirical estimates in Fig. 3, the non-stationary pattern of the spatialized XGumbel copula in Fig. 10 is generally reasonable. In contrast, the Brown–Resnick process in Fig. 11 not only fails to exhibit any non-stationarity, as expected since it is not designed to account for it, but it also yields rather low values with little spatial variability compared to the empirical estimates in Fig. 3.

5.3. Complete spatial models

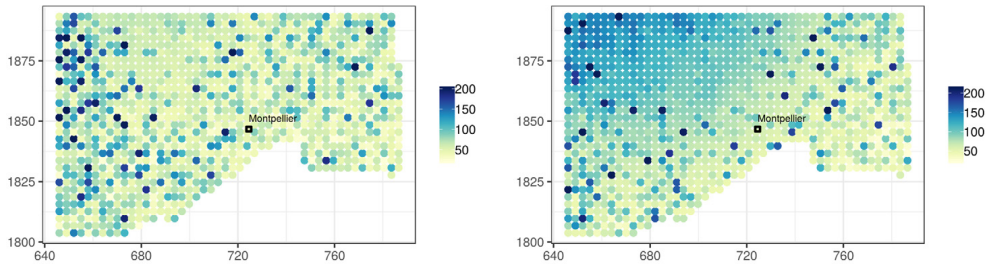
Simulations from the two complete fitted spatial models for maxima are illustrated in Figs. 12 and 13. In the former case, the dependence structure is modeled by the spatialized XGumbel copula whereas in the latter case, it is modeled by the Brown–Resnick process. In both cases, univariate marginal distributions are provided by the response surfaces for the GEV parameters from

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.



(a) *The reference station i is near Montpellier.* (b) *The reference station i is on the mountain range.*

Fig. 11. Maps of Brown–Resnick upper tail dependence coefficient estimates $\hat{\chi}_{ij}$, obtained by estimating the madogram with a sample of size 1000 (see Eq. (7)). The reference station i is shown by a white cross. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) (b)

Fig. 12. Two data-scale simulations of the spatialized XGumbel copula combined with the response surfaces for the GEV over the study area. The color scale is capped at the 99% quantile of the simulated values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Section 5.1. In the spatialized XGumbel copula case, the best set of parameters from the posterior distribution of the ABC scheme is used. The location of the disk center of the extra-parameter mapping, see Eq. (12), on the mountain range can easily be detected in Fig. 12. In the Brown–Resnick case, the grid for the simulation is restricted to two sub-areas (a first one encompassing the city of Montpellier and a second one in the mountain area) owing to computing limitations (Ribatet, 2018).

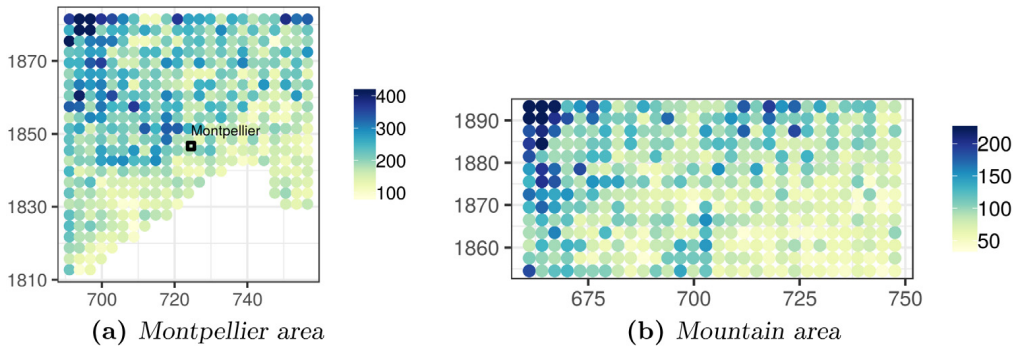
We then compare the two complete fitted spatial models for maxima in terms of a quantity that could be useful for regional hazard analysis. This quantity is related to the multivariate extension of the upper tail dependence coefficient termed m-dimensional joint tail dependence coefficients (Wadsworth and Tawn, 2013). Higher dimensional properties of the models can be investigated as these coefficients involve m-dimensional distributions instead of being limited to bivariate marginals as is the case for the upper tail dependence coefficient.

More precisely, we focus on trivariate properties, i.e. $m = 3$. Let Y_k, Y_i and Y_j represent annual maxima at three sites k, i and j respectively. Moreover, let R_k^T, R_i^T and R_j^T be the T-year return level at each site. The quantity of interest for our regional hazard analysis is the 3-dimensional joint tail dependence coefficient that is defined as follows:

$$\mathbb{P}(Y_i > R_i^T, Y_j > R_j^T | Y_k > R_k^T). \tag{14}$$

Note that, given that the univariate marginals are the same in both spatial models, differences in terms of the coefficient in Eq. (14) are only caused by differences in the spatial dependence

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.



(a) *Montpellier area* (b) *Mountain area*

Fig. 13. Two data-scale simulations of the Brown–Resnick process combined with the response surfaces for the GEV over the study area. The color scale is capped at the 99% quantile of the simulated values. Two sub-areas are selected as the implementation of the Brown–Resnick process we used did not allow simulation on the full area (Ribatet, 2018). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

structure. The interpolation ability of the spatial models is evaluated by setting the conditioning site k in Eq. (14) as one of the six validation stations not used for model inference (see the stations depicted with red filled circles in Fig. 2). For the other two sites i and j in Eq. (14), we selected two nearby sites from the calibration stations within a 20 km radius with the most complete observation record. These calibration stations wear numbers from 1 to 11 in Fig. 2.

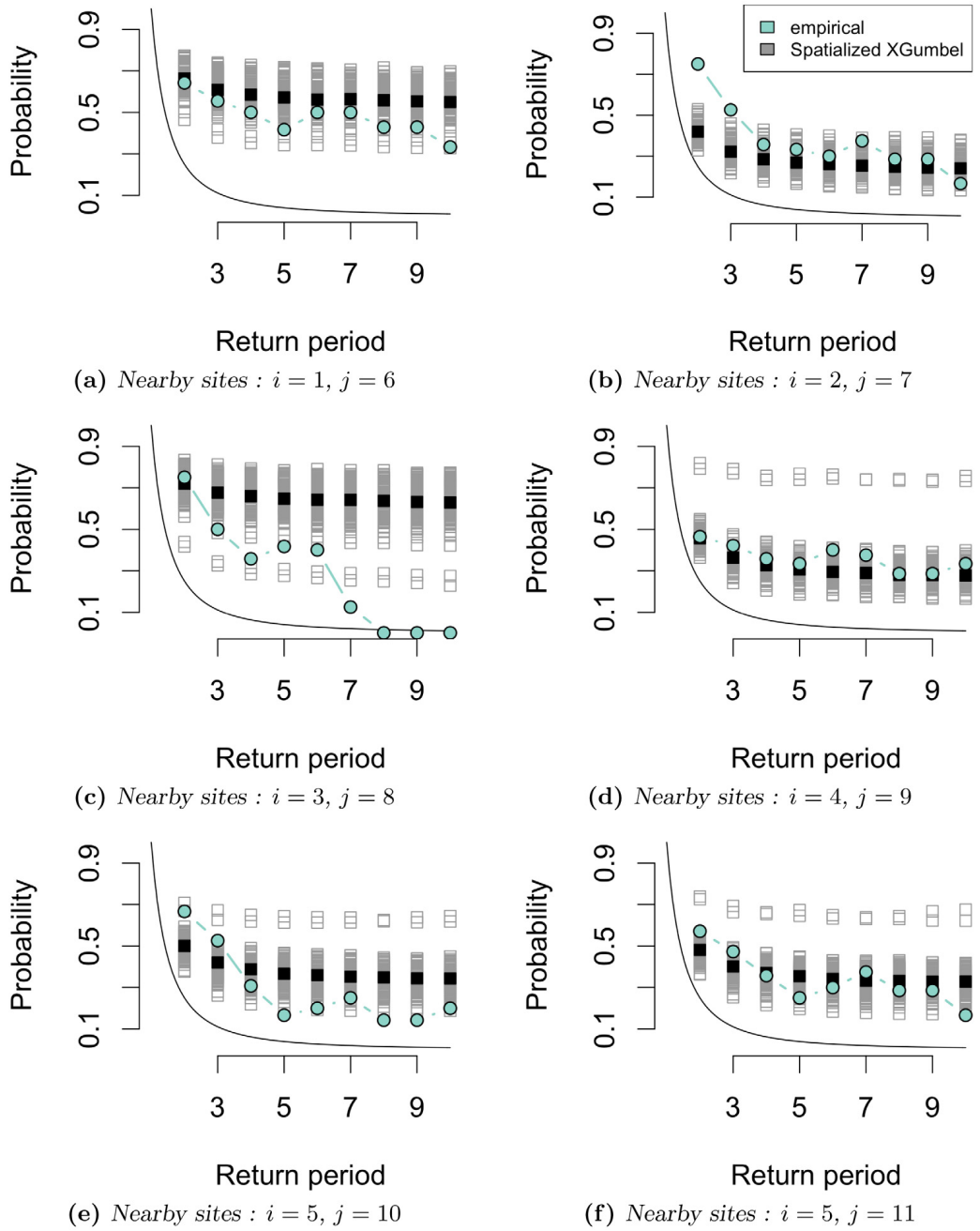
In Figs. 14 and 15, empirical and theoretical estimates of the 3-dimensional joint tail dependence coefficient from Eq. (14) are compared, with each of the six validation stations taken as the conditioning site k in turn. Empirical estimates, colored in light blue in both cases, are obtained by computing the sample proportions from the observed annual maxima with return levels determined from empirical quantiles. As there is no closed-form expression for Eq. (14), theoretical estimates are also deduced from proportions of samples of size 10000 simulated from each of the spatial models (GEV margins combined with either the spatialized XGumbel copula, in Fig. 14, or the Brown–Resnick process, in Fig. 15), with the return levels provided by the response surfaces for the GEV parameters (see Section 4.1, Eqs. (9)–(11)). For each return level, there are 100 theoretical estimates corresponding to different sets of parameters (from the posterior distribution resulting from the ABC scheme for the spatialized XGumbel copula or from the non-parametric bootstrap for the Brown–Resnick process). In addition, for the spatialized XGumbel copula, the median of the theoretical estimates of the 3-dimensional joint tail dependence coefficient is shown in black in Fig. 14 while, for the Brown–Resnick process, the theoretical estimates of the fit on the original calibration data are shown in black in Fig. 15.

As the dependence structure in both spatial models is max-stable, both theoretical coefficient estimates stabilize at longer return periods (greater than five years). Being a stationary model, the Brown–Resnick process always yields estimates at about the same level, wherever is located the conditioning site. In contrast, the spatialized XGumbel copula, thanks to its non-stationarity, can adapt to the location of the conditioning site. For instance, the estimates stabilize at about 0.5 for the validation station labeled "a" in Fig. 14(a) whereas they stabilize at about 0.3 for the validation station labeled "b" in Fig. 14(b). The empirical estimates are mostly contained within the spread of the theoretical estimates for both models, although it happens in a few instances that they fall outside, e.g. in Fig. 14(c) or Figs. 15(c) and 15(e). For some conditioning sites, e.g. Fig. 14(d), the spatialized XGumbel copula yielded two estimates that are far away from the others. These correspond to parameter vectors for which the disk centers are located near the coastline, see Fig. 8(c).

6. Conclusion

We proposed a spatial extension of the XGumbel copula that relies on the definition of the extra-parameters as a mapping of geographical covariates. Although the XGumbel copula could in

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.



(a) *Nearby sites :* $i = 1$, $j = 6$

(b) *Nearby sites :* $i = 2$, $j = 7$

(c) *Nearby sites :* $i = 3$, $j = 8$

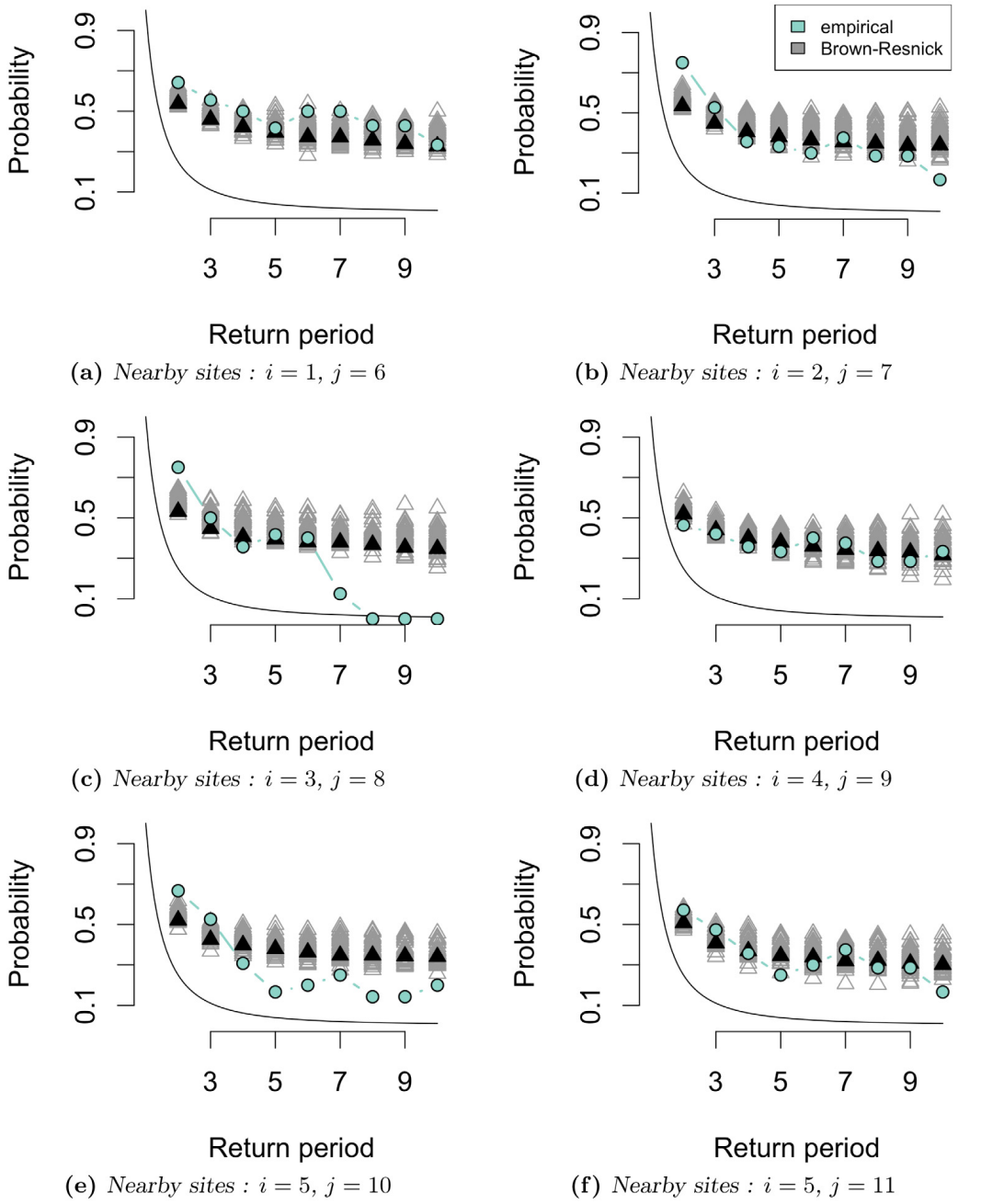
(d) *Nearby sites :* $i = 4$, $j = 9$

(e) *Nearby sites :* $i = 5$, $j = 10$

(f) *Nearby sites :* $i = 5$, $j = 11$

Fig. 14. 3-dimensional joint tail dependence coefficient estimates, see Eq. (14), with respect to return periods T on the x-axis. The Spatialized XGumbel estimates (gray squares) are proportions of simulated samples of size 10000 for each of the 100 sets of parameters of the posterior distribution. The median estimates are shown as black squares. The conditioning site k is one of the six validation stations, red filled circles in Fig. 2 wearing the letter corresponding to the panel. The other two sites i and j are calibration stations wearing numbers in Fig. 2 that are reported under each panel. The black line corresponds to the perfect independence case. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.



(a) *Nearby sites :* $i = 1,\ j = 6$ **(b)** *Nearby sites :* $i = 2,\ j = 7$

(c) *Nearby sites :* $i = 3,\ j = 8$ **(d)** *Nearby sites :* $i = 4,\ j = 9$

(e) *Nearby sites :* $i = 5,\ j = 10$ **(f)** *Nearby sites :* $i = 5,\ j = 11$

Fig. 15. 3-dimensional joint tail dependence coefficient estimates, see Eq. (14), with respect to return periods T on the x-axis. The Brown–Resnick estimates (gray triangles) are proportions of simulated samples of size 10000 for each of the 100 sets of parameters of the non-parametric bootstrap. The estimates from the fit on the original data are shown as black triangles. The conditioning site k is one of the six validation stations, red filled circles in Fig. 2 wearing the letter corresponding to the panel. The other two sites i and j are calibration stations wearing numbers in Fig. 2 that are reported under each panel. The black line corresponds to the perfect independence case. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

principle be fitted in high dimension, the large number of extra-parameters, corresponding to the number of sites in a spatial application, might hamper inference. The spatialized XGumbel copula is more parsimonious as it requires only a four parameter vector $\psi_{\text{spat}} = (\beta_B, \delta, \mu_x, \mu_y)$, independently of the number of sites. We designed the extra-parameter mapping shaped as a disk by relating the behavior of the strength of dependence between two sites, as characterized by the upper tail dependence coefficient χ, to desirable spatial properties. In particular, we focused on the pattern of decrease of the dependence with the distance by using χ estimates for five distance classes. These distance class χ estimates also serve as summary statistics in an ABC scheme to infer the parameters of the spatialized XGumbel copula. The spatialized XGumbel copula, when combined with a spatial regression model for the GEV marginal distributions, yields well-defined MEV distributions for any set of sites. Therefore, simulation is possible everywhere in the study area.

The proposed spatialized XGumbel copula is evaluated and compared with a Brown–Resnick process on annual maxima of daily precipitation totals in a region of the French Mediterranean with 177 gauged stations, six of which are kept for validation purposes. A vector generalized linear (VGLM) model is considered for the interpolation of the GEV parameters to model the univariate marginal distributions. The goodness-of-fit of the VGLM model is evaluated in terms of return levels at the validation stations. We analyzed the posterior distribution of the spatialized XGumbel copula parameters resulting from the rejection ABC scheme. Except for the parameter β_B, a global parameter inherited from one of the Gumbel copulas of the XGumbel, the ABC scheme inferred interpretable parameters. The Brown–Resnick process is fitted by pairwise log-likelihood minimization and uncertainty estimates are obtained by performing the fit on bootstrap resamples.

Comparison between the spatialized XGumbel copula and the Brown–Resnick process shows the following. The pattern of decrease of the strength of dependence is well reproduced in both cases. Owing to asymptotic dependence, the strength of dependence remains constant at extreme levels. However, strong non-stationarity patterns in the strength of dependence are present for the spatialized XGumbel copula whereas the Brown–Resnick process, by construction, has none. Simulations from both complete spatial models for maxima, GEV marginals together with spatial dependence structure, were provided for illustrations. A further downside of the Brown–Resnick process is that simulation on the full grid covering the study area was not possible due to computing limitations. We proposed a regional hazard analysis based on 3-dimensional joint tail dependence coefficients. These involve the trivariate distributions at three stations one of which is taken as a validation station and the other two are neighbor calibration stations. The simulations and the regional hazard analysis also highlight differences due to the presence or absence of non-stationarity in the dependence structures.

Earlier propositions to extend copulas to the spatial framework are based on a parametrization in terms of distance but are not especially targeting spatial maxima (Bárdossy and Li, 2008; Gräler, 2014; Krupskii et al., 2018). The construction of the XGumbel copula as the maximum between two weighted random variables is directly related to the max-mixture model (Wadsworth and Tawn, 2012; Bacro et al., 2016). Instead of relying on processes with well-defined spatial dependence structures, the spatial dependence of the spatialized XGumbel copula is driven by the mapping of extra-parameters. The shape of the mapping determines the non-stationarity pattern of the dependence structure. A completely different proposition to introduce non-stationarity in the dependence structure for spatial maxima was put forward in Huser and Genton (2016) concerning max-stable processes.

Further analyses are needed to develop and test different shapes for the extra-parameter mapping. An interesting development, that was already considered in preliminary work, would be to let the shape of the mapping change from year to year, leading to a conditionally max-stable model. This would allow, in particular, to let the areas of stronger and weaker dependence vary from one year to another. Another way to achieve this, while keeping the max-stable property, would be to iterate the extra-parametrization (or maximization) operation in Section 2.1, e.g. by assuming that the phenomenon of interest can be modeled as:

$$\max\{[\max(\boldsymbol{U}^{1/\boldsymbol{b}}, \boldsymbol{W}^{1/(\mathbf{1}-\boldsymbol{b})})]^{1/\boldsymbol{a}}, \boldsymbol{V}^{1/(\mathbf{1}-\boldsymbol{a})}\},$$

with $\boldsymbol{U} \sim C_{\beta_A}$, $\boldsymbol{V} \sim C_{\beta_B}$, $\boldsymbol{W} \sim C_{\beta_C}$, $\beta_A, \beta_B, \beta_C \geq 1$ are Gumbel copula parameters and $\boldsymbol{a}, \boldsymbol{b} \in [0, 1]^d$ two extra-parameter vectors. Although pairwise log-likelihood inference is widely

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

used, ABC inference scheme yields promising results. For complex dependence structure models, even pairwise log-likelihood might be intractable. We have used summary statistics that convey information on the strength of extremal dependence. Other statistics, for instance, conveying information on asymmetry or non-stationarity, as well as other ways to compute distances, such as the Wasserstein distance, could be considered (Arbel et al., 2019).

Acknowledgments

This work was supported by the French national program LEFE/INSU (CERISE and FRAISE). European programs PRIMA and ERANET-MEd (ALTOS and CHAAMS) Bi-national program Hubert-Curien, France (AMANDE).

References

Arbel, J., Crispino, M., Girard, S., 2019. Dependence properties and Bayesian inference for asymmetric multivariate copulas. J. Multivariate Anal. 174, 104530:1–20.
Bacro, J.-N., Gaetan, C., Toulemonde, G., 2016. A flexible dependence model for spatial extremes. J. Statist. Plann. Inference 172, 36–52.
Bárdossy, A., Li, J., 2008. Geostatistical interpolation using copulas. Water Resour. Res. 44, 1–15.
Beaumont, M.A., 2010. Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41, 379–406.
Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J., De Waal, D., Ferro, C., 2004. Statistics of Extremes Theory and Applications. John Wiley & Sons.
Benichou, P., Le Breton, O., 1987. Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. Une application de la méthode Aurelhy: la cartographie nationale de champs de normales pluviométriques. Météorologie.
Blanchet, J., Creutin, J.-D., 2017. Co-occurrence of extreme daily rainfall in the French Mediterranean region. Water Resour. Res. 53, 9330–9349.
Blanchet, J., Davison, A.C., 2011. Spatial modeling of extreme snow depth. Ann. Appl. Stat. 5, 1699–1725.
Braud, I., Ayral, P.-A., Bouvier, C., Branger, F., Delrieu, G., Le Coz, J., Nord, G., Vandervaere, J.-P., Anquetin, S., Adamovic, M., Andrieu, J., Batiot, C., Boudevillain, B., Brunet, P., Carreau, J., Confoland, A., Didon-Lescot, J.-F., Domergue, J.-M., Douvinet, J., Dramais, G., Freydier, R., Gérard, S., Huza, J., Leblois, E., Le Bourgeois, O., Le Boursicaud, R., Marchand, P., Martin, P., Nottale, L., Patris, N., Renard, B., Seidel, J.-L., Taupin, J.-D., Vannier, O., Vincendon, B., Wijbrans, A., 2014. Multi-scale hydrometeorological observation and modelling for flash flood understanding. Hydrol. Earth Syst. Sci. 18, 3733–3761.
Brown, B., Resnick, S., 1977. Extremes values of independent stochastic processes. J. Appl. Probab. 14, 732–739.
Brunet, P., Bouvier, C., Neppel, L., 2018. Retour d'expérience sur les crues des 6 et 7 Octobre 2014 à Montpellier-Grabels (Hérault, France): caractéristiques hydro-météorologiques et contexte historique de l'épisode. Géographie Phys. Et Environ. 12, 43–59.
Carreau, J., Naveau, P., Neppel, L., 2017. Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. Water Resour. Res. 53, 4407–4426.
Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. In: Springer Series in Statistics, Springer.
Coles, S.G., Heffernan, J.E., Tawn, J.A., 1999. Dependence measures for extremes value analyses. Extremes 2, 339–365.
Cooley, D., Naveau, P., Poncet, P., 2006. Variograms for spatial max-stable random fields. In: Dependence in Probability and Statistics. In: Lect. Notes Stat, vol. 187, Springer, New York, pp. 373–390.
Davison, A.C., Gholamrezaee, M.M., 2012. Geostatistics of extremes. Proc. R. Soc. Lond. Ser. A 468, 581–608.
Davison, A.C., Padoan, S.A., Ribatet, M., 2012. Statistical modeling of spatial extremes. Stat. Sci. 27, 161–186.
Delrieu, G., Nicol, J., Yates, E., Kirstetter, P.-E., Creutin, J.-D., Anquetin, S., Obled, C., Saulnier, G.-M., Ducrocq, V., Gaume, E., Payrastre, O., Andrieu, H., Ayral, P.-A., Bouvier, C., Neppel, L., Livet, M., Lang, M., du Châtelet, J.P., Walpersdorf, A., Wobrock, W., 2005. The catastrophic flash-flood event of 8-9 September 2002 in the Gard region, France: A first case study for the Cévennes-Vivarais Mediterranean hydrometeorological observatory. J. Hydrometeorol. 6, 34–52.
Durante, F., Salvadori, G., 2010. On the construction of multivariate extreme value models via copulas. Environmetrics 21 (2), 143–161.
Erhardt, R.J., Sisson, S.A., 2016. Modelling extremes using approximate Bayesian computation. In: Extreme Value Modeling and Risk Analysis. CRC Press, Boca Raton, FL, pp. 281–306.
Erhardt, R.J., Smith, R.L., 2012. Approximate Bayesian computing for spatial extremes. Comput. Statist. Data Anal. 56, 1468–1481.
Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency of the largest or smallest member of a sample. Math. Proc. Cambridge Philos. Soc. 24, 180–190.
Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22.
Genest, C., Favre, A.-C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. J. Hydrol. Eng. 12, 347–368.

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

Gnedenko, B.V., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math. 44, 423–453.
Gräler, B., 2014. Modelling skewed spatial random fields through the spatial vine copula. Spatial Stat. 10, 87–102.
Gumbel, E.J., 1958. Statistics of Extremes. Columbia Univ. Press, New York.
de Haan, L., 1984. A spectral representation for max-stable processes. Ann. Probab. 12, 1194–1204.
Huser, R., Genton, M.G., 2016. Non-stationary dependence structures for spatial extremes. J. Agric. Biol. Environ. Stat. 21 (3), 470–491.
Krupskii, P., Huser, R., Genton, M., 2018. Factor copula models for replicated spatial data. J. Amer. Statist. Assoc. 113, 467–479.
Lee, Xing Ju, Hainy, Markus, McKeone, James P., Drovandi, Christopher C., Pettitt, Anthony N., 2018. ABC model selection for spatial extremes models applied to South Australian maximum temperature data. Comput. Statist. Data Anal. 128, 128–144.
Liebscher, E., 2008. Construction of asymmetric multivariate copulas. J. Multivariate Anal. 99 (10), 2234–2250.
Marcon, G., Padoan, S.A., Naveau, P., Muliere, P., Segers, J., 2017. Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. J. Statist. Plann. Inference 183, 1–17.
Pickands, J., 1981. Multivariate extreme value distributions. In: Proceedings of the 43rd session of the International Statistical Institute, Bulletin de l'Institut International de Statistique, vol. 49, pp. 859–878, 894–902.
Ribatet, M., 2018. SpatialExtremes: Modelling Spatial Extremes. R package version 2.0-7.
Salvadori, G., De Michele, C., 2010. Multivariate multiparameter extreme value models and return periods: A copula approach. Water Resour. Res. 46, 1–11.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461–464.
Sibuya, M., 1960. Bivariate extreme statistics. Ann. Inst. Statist. Math. 11, 195–210.
Thibaud, E., Mutzner, R., Davison, A.C., 2013. Threshold modeling of extreme spatial rainfall. Water Resour. Res. 49, 4633–4644.
Vannitsem, S., Naveau, P., 2007. Spatial dependences among precipitation maxima over Belgium. Nonlinear Process. Geophys. 14, 621–630.
Wadsworth, J.L., Tawn, J.A., 2012. Dependence modelling for spatial extremes. Biometrika 99, 253–272.
Wadsworth, J.L., Tawn, J.A., 2013. A new representation for multivariate tail probabilities. Bernoulli 19, 2689–2714.
Yee, T.W., 2015. Vector Generalized Linear and Additive Models: With an Implementation in R. Springer, New York, USA.
Yee, T.W., Stephenson, A.G., 2007. Vector generalized linear and additive extreme value models. Extremes 10, 1–19.

Please cite this article as: J. Carreau and G. Toulemonde, Extra-parametrized extreme value copula : Extension to a spatial framework. Spatial Statistics (2020) 100410, https://doi.org/10.1016/j.spasta.2020.100410.

Annexe E

G20 - Generalized Pareto processes for simulating space-time extreme events : an application to precipitation reanalyses (2019).

Generalized Pareto processes for simulating space-time extreme events: an application to precipitation reanalyses

Fátima Palacios-Rodríguez, Gwladys Toulemonde, Julie Carreau, Thomas Opitz

Generalized Pareto processes for simulating space-time extreme events: an application to precipitation reanalyses

F. Palacios-Rodríguez [*†] G. Toulemonde[‡] J. Carreau [§] T. Opitz [¶]

Monday 16th December, 2019

Abstract

To better manage the risks of destructive natural disasters, impact models can be fed with simulations of extreme scenarios to study the sensitivity to temporal and spatial variability. We propose a semi-parametric stochastic framework that enables simulation of realistic spatio-temporal extreme fields using a moderate number of observed extreme space-time episodes to generate an unlimited number of extreme scenarios of any magnitude. Our framework draws sound theoretical justification from extreme value theory, building on generalized Pareto limit processes. For illustration on hourly gridded precipitation data in Mediterranean France, we calculate risk measures using extreme event simulations for yet unobserved magnitudes.

Keywords: extreme-value theory; precipitation; space-time Pareto processes; stochastic simulation; risk analysis.

1 Introduction

Extreme events of geophysical processes such as precipitation extend over space and time, and they can entail devastating consequences for human societies and ecosystems. Flash floods in southern France constitute highly destructive natural phenomena causing material damage and threatening human lives (Vinet et al., 2016), such as the two catasthrophic flash-flood events in Gard region on Septembre 2002 (Delrieu et al., 2005), and in Montpellier-Grabels on October 2014 (Brunet et al., 2018). Since damage and costs of floods have been increasing over the last decades, the understanding of temporal and spatial variability of rainfall patterns generating such floods receives considerable attention from the authorities (European Environment Agency, 2007). To help with this understanding, we develop a method to stochastically simulate realistic spatio-temporal extreme scenarios, which can be fed to impact models. Examples of impact models are urban flood models (such as shallow water models in Guinot and Soares-Frazão (2006) and Guinot et al. (2017)), which produce hydrological variables (such as water height or water speed), based on which experts make decisions about flood risk.

[*]Departamento de Estadística e Investigación Operativa, Facultad de Ciencias Matemáticas, Universidad Complutense de Madrid, Madrid, Spain.

[†]*Correspondence to: F. Palacios-Rodríguez, Departamento de Estadística e Investigación Operativa, Facultad de Ciencias Matemáticas, Universidad Complutense de Madrid, Plaza de Ciencias número 3, Madrid, 28040, Spain. Telephone: +34 91 394 4432. E-mail: fatima.palacios@ucm.es.*

[‡]IMAG, Université de Montpellier. CNRS. Inria. Montpellier. France.

[§]HydroSciences Montpellier, CNRS/IRD, Université de Montpellier. Montpellier. France.

[¶]Biostatistics and Spatial Processes, INRA Avignon. France.

Extreme-value theory (EVT) for spatial data proposes data-based stochastic modeling of such extreme events for predicting probabilities, risks and uncertainty behavior (Coles, 2001; de Haan and Ferreira, 2006; Ferreira and de Haan, 2014). Due to very complex deterministic and probabilistic patterns in such processes and the high dimension of data sets, realistic spatio-temporal modeling is challenging. In this work, we instead develop a data-driven non-parametric approach to handle extremal space-time dependence by transforming observed marginal quantiles in a spatially and temporally coherent way. We illustrate our method on a high-dimensional data set of gridded hourly reanalysis data. Our procedure draws sound justification from asymptotic theory for threshold exceedances with a strong probabilistic interpretation. We will explain how it allows us to flexibly define extreme episodes in space-time data based on different ways of aggregating marginal return periods over space and time.

Block-maxima and peaks-over-threshold (POT) methods are two widely known strategies in univariate EVT to identify extreme events in a data set. While the block-maxima method is based on the division of the observation period into non-overlapping periods of equal size (for instance months or years) to extract the maximum observation in each period (Ferreira and de Haan, 2015), the POT method consists in the study of positive exceedances above a given high threshold (Pickands III, 1975; Embrechts et al., 1997; Beirlant et al., 2004). Max-stable processes, introduced by de Haan (1984), are the natural infinite-dimensional generalization of the univariate generalized extreme value (GEV) distribution, which constitutes the only limiting distribution in block-maxima approach. Ferreira and de Haan (2014) and Dombry and Ribatet (2015) showed that generalized Pareto (GP) processes are the only possible asymptotic limits for threshold exceedances. Both approaches are closely linked through theoretical tail stability properties.

Several approaches were developed for stochastic simulation of spatial max-stable fields (Dombry et al. (2013, 2016); Oesting et al. (2018b,a)). Since max-stable processes are linked to the block-maxima approach, their realizations aggregate information of several of the underlying original events which may limit the physical interpretations of the simulated processes. Consequently, these simulations appear to be more appropriate for studying long-term events such as the erosion of the coastline (Chailan et al., 2017).

On the other hand, GP processes represent the original events that fulfill a threshold exceedance condition. They can be represented constructively by multiplying a random scaling variable with a so-called spectral process, the latter characterizing the spatial variation in the extreme events (Ferreira and de Haan (2014); Dombry and Ribatet (2015); Thibaud and Opitz (2015)). In practice, one usually first fits a parametric model for the spectral processes, and the estimators are then plugged in for simulation. In contrast, we here develop an algorithm for extracting observed spectral processes from data, and we then combine them with newly sampled scaling variables to generate new realizations of the extreme events.

Since extreme events are frequently spatio-temporal in nature, their extension and duration have to be accounted for. A semi-parametric method to simulate extreme spatio-temporal fields of wave heights in the Gulf of Lion (France) was proposed in Chailan et al. (2017) based on methods for the spatial setting developed by Caires et al. (2011) and Ferreira and de Haan (2014). The approach proposed in our work is motivated by Chailan et al. (2017) and provides three major novelties. Firstly, our procedure allows for an infinite number of simulations. Secondly, we embed our semiparametric resampling scheme in the framework of GP processes, which allows for a clear probabilistic interpretation of extreme events. Thirdly, a flexible general procedure is presented to identify extreme events and quantify their magnitude by accounting for space-time aggregation through homogeneous cost functionals that encapsulate operations such as averaging or taking maxima. In multivariate extreme value analysis, i.e., when the

observation domain consists of only a few points, our approach is closely related to empirical spectral measures, which have become a standard tool for estimating extremal dependence (e.g., Beirlant et al. (2004)).

The paper is structured as follows. Section 2 presents the theory for space-time GP processes. Techniques to practically implement and validate the spatio-temporal GP framework are proposed in Section 3. Our algorithm to generate extreme space-time scenarios is developed in Section 4. We illustrate our approach on hourly rainfall reanalysis data available on a 1 km^2 grid in Southern France over a 10-year period from 1997 to 2007 in Section 5. In this case study, we perform a comparative analysis based on two conventional risk measures using simulated extreme scenarios. Conclusions and future research are given in Section 6.

2 Theory of space-time GP processes

We write $\mathcal{S}$ for a compact subset of $\mathbb{R}^d$ to denote the area of interest and $\mathcal{T}$ for a compact subset of $\mathbb{R}^+$ to denote the time dimension, and we denote by $C(\mathcal{S}\times\mathcal{T})$ the space of continuous functions on $\mathcal{S}\times\mathcal{T}$, equipped with the supremum norm. The restriction of $C(\mathcal{S}\times\mathcal{T})$ to non-negative functions is written $C_+(\mathcal{S}\times\mathcal{T})$. Similarly, we define the space of non-negative continuous functions in $\mathcal{S}$ as $C_+(\mathcal{S})$.

In multivariate EVT, a GP limit was introduced in Rootzén and Tajvidi (2006) by conditioning on an exceedance event in at least one component. The aforementioned idea was extended to infinite-dimensional spaces by the definition of GP process in Ferreira and de Haan (2014) where the condition is based on exceedances of the supremum over space. To gain flexibility in the definition of the conditioning extreme events, Dombry and Ribatet (2015) provided the notion of ℓ-Pareto processes by considering more general exceedances defined in terms of a homogeneous cost functional denoted ℓ. Our focus here is on the spatial and temporal dimensions for the extent of extreme events. Since we aim to model phenomena that exceed a certain extreme threshold, we start by defining and characterizing space-time generalized ℓ-Pareto processes. The following constructive definition generalizes Dombry and Ribatet (2015).

2.1 Construction of space-time GP processes

We define a *cost functional* $\ell : C_+(\mathcal{S}\times\mathcal{T}) \to [0,+\infty)$ as a continuous nonnegative function that is homogeneous, i.e. $\ell(tf) = t\ell(f)$ for $t \geq 0$. Examples of such ℓ are the functions of maximum, minimum, average, or the value at a specific point $(s_0,t_0) \in \mathcal{S}\times\mathcal{T}$.

Definition 2.1 (Standard space-time ℓ-Pareto process)**.** *Let $W^* = \{W^*(s,t)\}_{s\in\mathcal{S},t\in\mathcal{T}}$ be a stochastic process in $C_+(\mathcal{S}\times\mathcal{T})$. We call W^* a standard space-time ℓ-Pareto process if it can be represented as*

$$W^*(s,t) \overset{d}{=} RY(s,t) \tag{1}$$

where

1. *Y is a stochastic process in $C_+(\mathcal{S}\times\mathcal{T})$ satisfying $\ell(Y) = 1$;*
2. *R has Pareto distribution with scale 1 and shape γ_R, i.e., $\mathbb{P}(R > r) = r^{-\gamma_R}$, $r > 1$;*
3. *Y and R are stochastically independent.*

The above definition is equivalent to the definition through the POT stability property: for any $u \geq 1$, the distribution of the renormalized threshold-exceeding process $\{u^{-1}W^* | \ell(W^*) \geq u\}$ is equal to the distribution of W^*; see Theorem 2 of Dombry and Ribatet (2015). By construction, we get $Y \stackrel{d}{=} W^*/\ell(W^*)$ and $R \stackrel{d}{=} \ell(W^*)$. A generalized version of such Pareto processes is given in Definition 2.2 by allowing for flexibility in the marginal distributions according to the location-scale-shape parametrization commonly used in univariate EVT.

Definition 2.2 (Generalized space-time ℓ-Pareto process)**.** *Given an ℓ-Pareto process $W^*(s,t)$ constructed according to Definition 2.1 and continuous real functions $\sigma(s,t) > 0$, $\mu(s,t)$ and $\gamma(s,t)$ in $C(\mathcal{S} \times \mathcal{T})$, a generalized space-time ℓ-Pareto process is any process constructed as*

$$W(s,t) \stackrel{d}{=} \begin{cases} \mu(s,t) + \sigma(s,t)\{W^*(s,t)^{\gamma(s,t)} - 1\}/\gamma(s,t), & \gamma(s,t) \neq 0, \\ \mu(s,t) + \sigma(s,t) \log W^*(s,t), & \gamma(s,t) = 0. \end{cases} \tag{2}$$

2.2 Asymptotic results for space-time GP processes

We shortly recall the two main asymptotic results for characterizing extremes of stochastic processes : max-stable processes and Pareto processes. We refer the reader to the literature for technical details (Lin and de Haan, 2001; de Haan and Ferreira, 2006; Ferreira and de Haan, 2014; Thibaud and Opitz, 2015; Dombry and Ribatet, 2015). We use the symbol "$\Rightarrow$" to represent variants of weak convergence of random elements from the univariate, multivariate or functional domain.

Consider independent copies $X_1, \ldots, X_n$ of a stochastic space-time process $X = \{X(s,t)\}_{s \in \mathcal{S}, t \in \mathcal{T}}$ with continuous trajectories. We say that the process X is in the functional maximum domain of attraction of a max-stable process $Z = \{Z(s,t)\}_{s \in \mathcal{S}, t \in \mathcal{T}}$ with continuous trajectories if there exists continuous functions $a_n > 0$ and b_n such that

$$\left\{ \max_{1 \leq i \leq n} \frac{X_i(s,t) - b_n(s,t)}{a_n(s,t)} \right\}_{s \in \mathcal{S}, t \in \mathcal{T}} \Rightarrow \{Z(s,t)\}_{s \in \mathcal{S}, t \in \mathcal{T}}. \tag{3}$$

Further details about space-time max-stable processes can be found in Davis et al. (2013a,b).

The convergence of the dependence structure and of marginal distributions in (3) can be studied separately; see de Haan and Ferreira (2006, Section 9.2). A standardised process $X^* = \{X^*(s,t)\}$ can be defined by $X^*(s,t) = H^{-1}(F_{(s,t)}(X(s,t)))$, $s \in \mathcal{S}$, $t \in \mathcal{T}$, where H^{-1} denotes the inverse function of the standard Pareto distribution function H, and $F_{(s,t)}$ denotes the distribution of $X(s,t)$. If X has continuous marginal distributions $F_{(s,t)}$, then X^* has marginal standard Pareto distributions. For $a_n \equiv n$, $b_n \equiv 0$, the max-stable limit for X^* in (3) is a standard max-stable process $Z^* = \{Z^*(s,t)\}_{s \in \mathcal{S}, t \in \mathcal{T}}$ with unit Fréchet marginal distributions; see de Haan and Ferreira (2006, Definition 9.2.4).

If X^* is in the maximum domain of attraction of a max-stable process Z^* and the cost functional ℓ is continuous at 0, we get the convergence of *ℓ-exceedances* on the standard scale:

$$\left\{ u^{-1} X^*(s,t) | \ell(X^*(s,t)) > u \right\} \Rightarrow \{W^*(s,t)\}, \; u \to \infty, \tag{4}$$

where $W^*(s,t)$ is a standard space-time ℓ-Pareto process as in Definition 2.1 (Dombry and Ribatet, 2015, Theorem 3). Conversely, if the convergence in (4) holds for ℓ chosen as the maximum norm, then convergence in (3) of the max-stable process X^* to Z^* follows. An example of Pareto processes with log-Gaussian profile process is given in Appendix A.

3 Practice of space-time GP processes

In practice, we use the asymptotic theory exposed in Section 2 for conducting statistical analyses on extreme events based on finite-sample data, which poses a number of practical challenges. In this section, we propose solutions for three issues: the standardisation of marginal distributions (Section 3.1), the definition of extreme space-time episodes (Section 3.2), the analysis and verification of asymptotic stability properties (so-called *threshold-stability*, see Section 3.3).

3.1 Marginal transformations

We first discuss suitable marginal transformations of X such that X^* satisfies convergence with respect to ℓ-exceedances in (4). In theory, values of $X^*(s,t)$ close to 0 are pushed to 0 when $u \to \infty$ in (4), but in practice the use of a high but finite threshold u leads to non-zero values in $u^{-1}X^*(s,t)$. Therefore, a certain ambiguity persists in practice to define the standardisation for relatively small, non extreme values of $X(s,t)$. In particular, if the minimum value of the data process X arises with positive and non negligible probability, such as the value 0 for the absence of precipitation in our application study, then this minimum value should be mapped to 0 in the standardised process X^*. Here, we develop the general idea of such transformations and a more specific transformation for precipitation data is proposed in Section 5. We choose a distribution function $G : \mathbb{R} \to [0,1]$ whose survival function $\bar{G}$ verifies: $x\,\bar{G}(x) \to 1,\ x \to \infty$, and $\bar{G}(0) = 1$; we write $G^{\leftarrow}$ for the (generalized) inverse function of G. We then define the transformation $T = T_{(s,t)} : \mathbb{R} \to [0,\infty)$ towards the standardised process X^* as follows:

$$X^*(s,t) = T(X(s,t)) = G^{\leftarrow}(F_{(s,t)}(X(s,t))) \tag{5}$$

where $F_{(s,t)} : \mathbb{R} \to [0,1]$ denotes the distribution of $X(s,t)$. The (generalized) inverse transformation of T can be defined as $T^{\leftarrow}(f) = F^{\leftarrow}_{(s,t)}(G(f))$ for $f \in C_+(\mathcal{S} \times \mathcal{T})$, with $F^{\leftarrow}_{(s,t)}$ the (generalized) inverse function of $F_{(s,t)}$.

Regarding marginal modeling, it is natural to use a tail representation motivated by univariate EVT, whose parametrization corresponds directly to the GP process in Definition 2.2. For a fixed high threshold function $u(s,t)$, we assume that

$$\mathbb{P}(X(s,t) > x) = 1 - F_{(s,t)}(x) = \left[1 + \gamma(s,t)\frac{x - \mu(s,t)}{\sigma(s,t)}\right]_+^{-1/\gamma(s,t)} \tag{6}$$

for $x > u(s,t)$, with parameter functions for position $\mu(s,t) < u(s,t)$, for scale $\sigma(s,t) > 0$ and for shape $\gamma(s,t)$, such that the right-hand side of (6) is less than 1 (Thibaud and Opitz, 2015). For data values $X(s,t)$ below $u(s,t)$, we may use appropriately chosen empirical distribution functions or any other useful model, where the probability mass below $u(s,t)$ should amount to $F_{(s,t)}(u(s,t))$ with $F_{(s,t)}$ defined in (6).

The standardisation in (5) leads to $\mathbb{P}(T(X(s,t)) > T(x)) \sim \frac{1}{T(x)}$ for large x, and therefore to $\mathbb{P}(T(X'(s,t)) > T(X(s,t)) \mid X(s,t) = x(s,t)) \sim \frac{1}{T(x(s,t))}$ for an independent copy X' of X. For the observed $X(s,t)$, the value of $T(X(s,t))$ can be interpreted as the (marginal) return period of the observation $X(s,t)$, and at high quantiles we can interpret X^* as the space-time process of marginal return periods. The cost functional ℓ (approximately) aggregates marginal return periods $X^*(s,t)$ into return periods $\ell(X^*)$ for space-time episodes. For details about the definition of return periods, see Section 5.7.

3.2 Defining extreme episodes

For the purpose of simulating realistic spatio-temporal extreme scenarios, we have to define what "extreme" means. With environmental data, we often have only a single observation of the space-time process X, and very high values typically tend to cluster temporally within relatively short sub-periods. We consider such sub-periods as extreme space-time events. If it is realistic to assume that temporal dependence of extremes becomes negligible for relatively large time lags, theoretical results based on independent processes as in Section 2 can be used. In the space-time GP process framework, the value of $\ell(X)$ quantifies the magnitude of events. In practice, we apply ℓ to a large collection of candidate episodes to extract the most extreme ones. Our extraction algorithm is designed to avoid temporal intersection of the selected extreme episodes.

There is no unique definition of an extreme event, i.e. of the cost functional ℓ, rather it depends on the nature of the considered phenomenon, on the data set, on the objective of the study, and also on the structure of the model (McPhillips et al., 2018). Expert knowledge may suggest how to measure the extreme nature of an event, where the question of how to combine criteria related to duration, spatial extent and magnitude is recurrent. For instance, French et al. (2018) develop new visualizations of extreme heat waves by composing a temporal and spatial cost functional. Chailan et al. (2017) extract extreme wave heights based on spatio-temporal maxima in sliding time windows.

In the following, we use the idea of sliding space-time windows and specify the support of the cost functional ℓ introduced in Section 2.1 as a *neighborhood* $\mathcal{N}(s,t)$ at location $s \in \mathcal{S}$ and at time $t \in \mathcal{T}$. In practice, the window size defines the maximal time duration and spatial extent of extreme events. The space index s may be missing if we consider the full study area for extracting extreme events. This neighborhood could be defined through an event duration δ in time, and the spatial support could be the full study area or a sub-region such as a catchment or a certain distance buffer around a specific site s_0. To indicate the local support of the cost functional defined as a neighborhood around (s,t), we use the notation $\ell_{s,t}(X^*) = \ell(\{X^*(s',t'),\ (s',t') \in \mathcal{N}(s,t)\})$.

We propose to define $\mathcal{N}(s,t)$ as the product of a spatial neighborhood $\mathcal{N}(s)$ (e.g., $\{s' \in \mathcal{S} \mid \|s-s'\| \leq 15 \text{ km}\}$) and a temporal neighborhood $\mathcal{N}(t)$ (e.g., $\{t' \in \mathcal{T} \mid |t-t'| \leq 12 \text{ hours}\}$), $\mathcal{N}(s,t) = \mathcal{N}(s) \times \mathcal{N}(t)$. The above choice of the spatial extension of the neighborhood and the time duration takes into account the spatial and temporal dependence of extreme episodes in our dataset, and can be seen as a local smoothing of the data. Useful cost functionals ℓ for space-time episodes are obtained by composing a spatial functional ℓ^S with a temporal functional ℓ^T, the latter applied to the values of ℓ^S observed over a number of consecutive time steps :

$$\ell_{s,t}(X^*) = \ell^T(\ell^S_{s,t-(\delta-1)}(X^*), \ldots, \ell^S_{s,t}(X^*)), \tag{7}$$

with $\ell^S_{s,t}(X^*) = \ell^S(\{X^*(s',t) \mid s' \in \mathcal{N}(s)\})$ and δ the duration of the episode. Moreover, based on $\ell^S_{s,t}(X^*)$ we can define cost functionals that combine the values obtained for all spatial neighborhoods $\mathcal{N}(s)$ by taking their maximum value (or again, any other spatial aggregation value). In this case, we define:

$$\ell_t(X^*) = \ell^T\left(\max_{s\in\mathcal{S}} \ell^S_{s,t-(\delta-1)}(X^*), \ldots, \max_{s\in\mathcal{S}} \ell^S_{s,t}(X^*)\right). \tag{8}$$

If X satisfies the functional domain of attraction condition (3), then

$$\mathbb{P}(\ell(X^*) > u) \sim \theta_\ell / u, \quad u \to \infty, \tag{9}$$

where θ_ℓ is the *ℓ-extremal coefficient* (for details, see Engelke et al., 2018). When $\ell_{s,t}$ corresponds to the maximum function over $\mathcal{N}(s,t)$ (i.e., $\ell^T = \max$ and $\ell^S_{s,t} = \max$), the ℓ-extremal coefficient $\theta_{\ell_{s,t}}$ defines the classical extremal coefficient of the domain $\mathcal{N}(s,t)$ (see Example 4 of Engelke et al., 2018).

Using (9), we can calculate approximate return levels for extreme episodes characterized as ℓ-exceedances above a large threshold u. The simplest case arises for $\theta_\ell = 1$, i.e., when θ_ℓ is known beforehand and we do not have to estimate it from data. For instance, if $(s_0, t_0) \in \mathcal{S} \times \mathcal{T}$ is a fixed space-time point, we can define the cost functional value $\ell(X^*)$ as $X^*(s_0, t_0)$, and $\theta_\ell = 1$. Moreover, $\theta_\ell = 1$ if ℓ is the average, i.e. $\ell_{s,t}(x) = \frac{1}{|\mathcal{N}(s,t)|} \int_{\mathcal{N}(s,t)} x(s',t') \, \mathrm{d}(s',t')$; see Ferreira et al. (2012, Proposition 2.2). When $\theta_\ell \neq 1$, an estimator of θ_ℓ can be plugged into (9), such as a weighted least square estimator (Engelke et al., 2018). Finally, since $\mathbb{P}(\ell((X')^*) > \ell(X^*) \mid X^* = x^*) \sim \theta_\ell / \ell(x^*)$ at high quantiles of $\ell(X^*)$ for an independent copy X' of X, we can interpret $\ell(x^*)/\theta_\ell$ as the return period of an extreme event x^*.

3.3 Techniques to analyze asymptotic dependence

The functional domain of attraction condition in (3) is the theoretical basis for using GP processes. It requires that a relatively strong type of extremal dependence, known as asymptotic dependence, prevails in the data-generating process X, at least for small distances in space and time. With asymptotic dependence between two points (s,t) and $(s',t') = (s + \Delta s, t + \Delta t)$, we observe a strictly positive limit of the probability $P\left(F_{(s',t')}(X(s',t')) > u \mid F_{(s,t)}(X(s,t)) > u\right)$ as $u \to 1$. In this case, so-called threshold stability holds when moving towards higher quantiles, such that the typical spatial and temporal extent of clusters of extreme values does not depend on event magnitude. In practice, we should verify that data exhibit such asymptotic dependence. We shortly discuss two approaches : the study of empirical extremal coefficients and, the assessment of the independence of observed scale variable $\ell(X^*)$ and profile process $X^*/\ell(X^*)$.

3.3.1 Spatial and temporal extremal coefficient functions

Pairwise extremal coefficients provide a summary of extremal dependence with respect to distance in space and time and are calculated from bivariate data; see Appendix B for details on empirical estimation. We consider first the spatial extremal coefficient function $\theta^{spa}(h)$ to measure extremal dependence between sites separated by spatial distance h at a given time, and second the temporal extremal coefficient function $\theta^{tim}(k)$ to measure extremal dependence for a time lag k at a given site. We estimate $\theta^{spa}(h)$ using observation pairs with structure $(X(s,t_i), X(s+\Delta s, t_i))$ where $\Delta s = h$, and we estimate $\theta^{tim}(k)$ from observation pairs $(\max_{s\in\mathcal{S}} X(s,t_i), \max_{s\in\mathcal{S}} X(s,t_i+k))$.

3.3.2 Independence of scale and profile

The POT stability manifests itself through the (approximate) independence between the profile process $Y = X^*/\ell(X^*)$ and the random scale $R = \ell(X^*)$ for $\ell(X^*) > u$. In practice, the threshold u should be high enough for this property to hold approximately, such that the limit process in (4) becomes a useful approximation to data. Due to the very high dimension of the profile process in the space-time setting, it is difficult to check this independence directly based on observed scales and profiles. Instead, we propose to check for the absence of strong trends in summary statistics of Y with respect to the event magnitude R, which would indicate dependence between Y and R.

In our application, we will focus on checking the scale-profile independence in space by considering the set of extreme spatial episodes W_t^* satisfying $\ell_t^S(W_t^*) > u$, and we use two summary statistics calculated from the profile processes $Y_t = W_t^*/\ell_t(W_t^*)$ in $C_+(\mathcal{S})$. First, we consider $f_{u'}(Y_t)$ defined as the proportion of sites s where $Y_t(s) \leq u'$: useful values of u' are relatively small or large quantiles of Y_t, to check for trends in the magnitude of Y_t with respect to $\ell_t(W_t^*)$. Second, we consider the empirical standard deviation $sd(Y_t')$ of $Y_t'(s) = \sqrt{Y_t(s)}$: if there are trends with respect to event magnitude, we usually find trends of $sd(Y_t')$. The square root transformation ensures finite standard deviation values.

Several empirical studies on climatic data show that extremal dependence may weaken when the event magnitude increases (Opitz et al., 2015; Huser and Wadsworth, 2018; Le et al., 2018; Tawn et al., 2018). Then, asymptotic independence may ultimately arise, or the dependence strength may stabilize at very high but unobserved magnitudes. We cannot check this stability behavior with absolute certainty in finite samples. If the extremal dependence strength continues to weaken in data above the selected threshold u, we acknowledge that the GP process framework leads to rather conservative probability estimates for observing concomitant high values.

4 Methodology for uplifting observed extreme episodes

We now describe the general procedure for the extraction of extreme space-time episodes (Section 4.1) and the algorithm to resample new space-time scenarios (Section 4.2). A probabilistic interpretation of such resampling scheme is given in Section 4.3. Throughout and without loss of generality, we here use the same notation for the single observation of the space-time process $X(s,t)$ and the stochastic process itself.

4.1 Selection of extreme episodes

Algorithm 1 describes the extraction of extreme episodes from standardised data X^*. To start, we define the space-time neighborhoods $\mathcal{N}(s,t)$ whose intensities are assessed by applying the cost functional ℓ. If the neighborhood is the full study region, we may drop the index s and simply write $\mathcal{N}(t)$. We choose a threshold u for the cost functional for which the asymptotic stability properties underpinning our approach are (approximately) satisfied. There must be at least one exceedance of the cost functional above the threshold in the data set. The first step of the algorithm is to compute the values of ℓ for each neighborhood $\mathcal{N}(s,t)$. We select as the first extreme episode the neighborhood $\mathcal{N}(s_1,t_1)$ where $\ell_{s,t}$ reaches its maximum value ℓ_1. We aim at extracting a collection of extreme episodes that are at most weakly dependent; therefore, the algorithm needs a mechanism to "decluster" extreme episodes. The second extracted extreme episode corresponds to the maximum value of $\ell_{s,t}(X^*)$ arising in the data set $X^*(s,t)$ with t in the set of reduced time steps after removal of time steps that intersect with $\mathcal{N}(s_1,t_1)$ or, more generally, with a larger temporal buffer zone $\mathcal{N}_{\text{buffer}}(t_1)$ around t_1 involving a buffer parameter $\beta \geq 0$ to remove more time steps. We then iterate this procedure of episode extraction and data set reduction. The stopping criterion for the extraction of extreme episodes is two-fold: either a fixed target number m' of extreme episodes is reached, or the extreme condition $\ell_{s,t}(X^*) > u$ for a fixed high threshold u cannot be fulfilled any longer in the reduced data set.

If the maximum of $\ell_{s,t}(X^*)$ is not unique and is realized at several coordinates (s,t), we must define a rule to extract a single (s,t) that identifies the corresponding extreme space-time episode. In particular, if we find several consecutive time steps t where $\ell_t(X^*)$ in Equation (8) is equal to the maximum, we fix

the anchor time step t of the extreme episode as follows. Usually, δ consecutive values are equal, and we then set t to the closest value below or equal to the median of these time steps. This rule will tend to center the extreme space-time episode on the strongest values in X^*. That is, if the maximum arises at time steps $t_0, \ldots, t_0 + \delta - 1$, we fix $t = t_0 + \lfloor \frac{\delta}{2} \rfloor$ as the anchor time step of the extreme space-time episode.

Algorithm 1: Algorithm for selecting extreme episodes defined over space-time neighborhoods $\mathcal{N}(s,t)$. In Step 8, instead of extracting only the extreme neighborhood $\mathcal{N}(s_i, t_i)$, we may sometimes want to extract the full study domain $\mathcal{N}(t_i) \times \mathcal{S}$.

Input:

- $\{X^*(s,t), s \in \mathcal{S}, t \in \mathcal{T}\}$, space-time observations on a standardised scale;
- $\mathcal{S}' \subseteq \mathcal{S}$ sites of interest and $\mathcal{T}' \subseteq \mathcal{T}$ time steps of interest.
- m' the maximum number of extreme episodes to select;
- u threshold on $\ell_{s,t}(X^*)$ for the selection of extreme episodes;
- $\delta > 0$ the duration of extreme episodes defining temporal neighborhoods $\mathcal{N}(t) = [t - (\delta - 1), t]$;
- $\beta \geq 0$ buffer time step to ensure independent extreme episodes defining extended temporal neighborhoods $\mathcal{N}_{\text{buffer}}(t) = [t - (\delta - 1) - \beta, t + (\delta - 1) + \beta]$;
- $\mathcal{N}(s)$ spatial neighborhood for $s \in \mathcal{S}'$, such that $\mathcal{N}(s,t) = \mathcal{N}(s) \times \mathcal{N}(t)$.

Output:

- m: the number of selected extreme episodes ($m \leq m'$);
- $\left\{X^*_{[1]}, X^*_{[2]}, \ldots, X^*_{[m]}\right\}$, $\{s_1, s_2, \ldots, s_m\}$, $\{t_1, t_2, \ldots, t_m\}$, $\{\ell_1, \ell_2, \ldots, \ell_m\}$: collection of extreme episodes; observation sites and times; aggregation values related to extreme episodes.

1 **begin**
2 Set $\mathcal{I} = \mathcal{T}'$.
3 Calculate $\ell_{s,t}(X^*)$ for all $t \in \mathcal{T}', s \in \mathcal{S}'$ with $\mathcal{N}(s,t) \subset \mathcal{S} \times \mathcal{T}$.
4 $i \leftarrow 1$.
5 **while** $i \leq m'$ *and* $\max_{s \in \mathcal{S}', t \in \mathcal{I}} \ell_{s,t}(X^*) > u$ **do**
6 $(s_i, t_i) \leftarrow \arg\max_{t \in \mathcal{I}, s \in \mathcal{S}'} \ell_{s,t}(X^*)$
7 $\ell_i \leftarrow \ell_{s_i,t_i}(X^*)$
8 $X^*_{[i]} \leftarrow \{X^*(s', t'),\ (s', t') \in \mathcal{N}(s_i, t_i)\}$
9 $\mathcal{I} \leftarrow \mathcal{I} \setminus \mathcal{N}_{\text{buffer}}(t_i)$
10 $i = i + 1$
11 **return** m, $\left\{X^*_{[1]}, X^*_{[2]}, \ldots, X^*_{[m]}\right\}$, $\{s_1, s_2, \ldots, s_m\}$, $\{\ t_1, t_2, \ldots, t_m\}$, $\{\ell_1, \ell_2, \ldots, \ell_m\}$

4.2 Semi-parametric simulation method

To sample new extreme space-time scenarios, we proceed as follows:

1. **Standardisation**: Estimate marginal tail parameter functions $\gamma(s,t)$, $\sigma(s,t)$ and $\mu(s,t)$ in (6), and denote by $X^* = \{T(X(s,t))\}_{s\in\mathcal{S},t\in\mathcal{T}}$ the resulting standardised process (5).

2. **Selection of extreme episodes**: Fix the maximum number of extreme episodes m'. Use Algorithm 1 to extract the collection of $m \leq m'$ extreme episodes $X^*_{[i]}$, $i = 1,\ldots,m$.

3. **Lifting**: Sample R_i, $i = 1,\ldots,m$ according to a Pareto distribution with shape 1 and scale $\alpha > 0$, i.e. $\mathbb{P}(R_i > x) = \alpha/x$, $x \in [\alpha,\infty)$, and generate lifted extreme episodes as

$$V_i(s,t) = R_i \frac{X^*_{[i]}(s,t)}{\ell_i} = R_i Y_i(s,t),\ (s,t) \in \mathcal{N}(s_i,t_i). \tag{10}$$

4. **Back-transformation to original scale**: Lifted extreme episodes are transformed back to the original marginal scale by $W_i(s,t) = T^{\leftarrow}(V_i(s,t))$, $(s,t) \in \mathcal{N}(s_i,t_i)$.

When fixing the value m' of the number of extreme episodes to extract, we aim for a representative sample of spatio-temporal extremal patterns in the data, but have to keep in mind that for a large value of m' the POT stability property may not be satisfied.

4.3 Interpretation of our proposed model

According to Definition 2.2, the lifting procedure in Section 4.2 samples new realizations V_i of a space-time Pareto process with support $\mathcal{N}(s_i,t_i)$ for each extreme episode i. Since $\mathbb{P}(\ell(X^*) > x) \sim \theta_\ell/x$ for large x and since resampled scale variables R_i are larger than α, we obtain α/θ_ℓ as the minimum return period for resampled extreme episodes. Moreover, choosing a larger α will generate resampled extreme episodes with longer return period. By the construction of the simulation method and the POT property, uplifted scenarios have the same spatial patterns of variability as observed values, but they correspond to longer return periods.

We can further establish a link between our resampling procedure and the linear normalization constants in (3) leading to a max-stable limit at the original marginal scale. A valid choice for b_n, as suggested by EVT, is the $(1-1/n)$-quantile of $F_{(s,t)}$. With resampled scaling variable $R_i = r_i$ and the originally observed one ℓ_i (see Sections 4.1-4.2), and with appropriately chosen events A, we can follow arguments similar to Chailan et al. (2017, Appendix) and show that

$$P\left(\frac{W_i(s,t)-b_{nc}}{a_{nc}} \in A \middle| R_i = r_i, \ell_{s,t}(X^*) = \ell_i\right) \approx P\left(\frac{X_{[i]}(s,t)-b_n}{a_n} \in A \middle| R_i = r_i, \ell_{s,t}(X^*) = \ell_i\right),$$

as $n \to \infty$, where $c = r_i/(\ell_i\theta_\ell)$. Therefore, the resampled and backtransformed episode $W_i(s,t)$ has approximately the same probability distribution as the observed extreme episodes in $X(s,t)$, except for b_n and a_n replaced by b_{nc} and a_{nc}. If $\alpha > \theta_\ell\ell_i$, then $b_{nc} > b_n$, and our procedure generates a threshold-stable stochastic process at a higher level than the observed one.

5 Application to precipitation in Mediterranean France

We use our resampling algorithm to produce large numbers of realistic spatio-temporal extreme precipitation scenarios in a region in Mediterranean France where flash floods are frequent. Furthermore, we show how to calculate two risk measures for the most extreme observed space-time episodes before and after uplifting them to longer return periods.

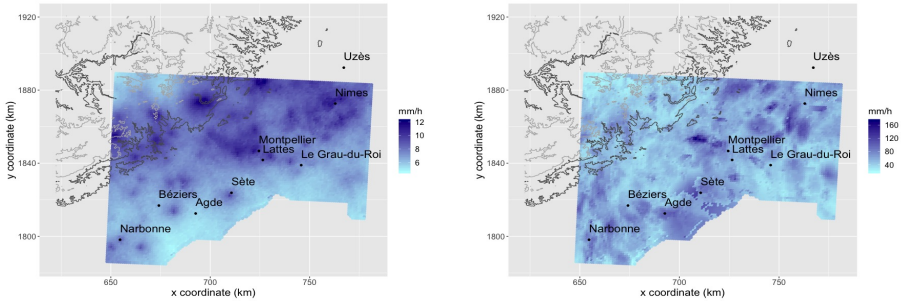


Figure 1: Empirical return levels at 98% level (left panel) and maxima (right panel) of hourly precipitation intensities for each grid cell in our study area from 1997 to 2007. Grey and black contour lines indicate altitude (400 and 800 m respectively).

5.1 Description of the data set

Our semi-parametric approach does not provide a mechanism to spatially interpolate observations. Therefore, precipitation measurements should be available over a sufficiently dense network of sites. We use hourly precipitation reanalysis data over a 1 km^2 grid, constructed by merging radar signals and observed hourly precipitation totals (Tabary et al., 2012). The grid has $10,914$ cells covering a 133.2 km $\times$ 104.3 km area in Mediterranean France, see Figure 1, with $87,642$ hourly time steps covering the 10-year period from 1997 to 2007. The unit of measurement is mm/h. This data set was provided by *Météo-France* (http://www.meteofrance.com). The large dimension of the data set allows us to disregard restrictive parametric assumptions in favour of a nonparametric approach for the extremal dependence model.

Empirical return levels of rainfall intensities at the 98% level (i.e., of strictly positive observations) and the maximum precipitation values observed over the complete study period are reported for each grid cell in Figure 1.

5.2 Standardisation of marginal distributions

The first step of our lifting procedure is the definition of a marginal transformation T, appropriate for extreme hourly precipitation data, to obtain the standardised process X^* in (5). We first discuss our choice of the target distribution G. Due to the hourly temporal resolution, zero values occur with very high frequency in the data. Therefore, we include a discrete mass p_0 at 0 to represent the absence of precipitation. Following Opitz (2016), we construct G to have a mass $p_0 \geq 0$ at 0, a uniform density on $(0, x_0)$, and a standard Pareto distribution for $x > x_0$ where $x_0 > 1$. The junction point x_0 is chosen to ensure the continuity of the density of G for $x > 0$:

$$G(x) = \begin{cases} 0, & x < 0, \\ p_0, & x = 0, \\ p_0 + \frac{(1-p_0)^2}{4} x, & 0 < x \leq 2/(1-p_0), \\ 1 - 1/x, & x > 2/(1-p_0). \end{cases} \tag{11}$$

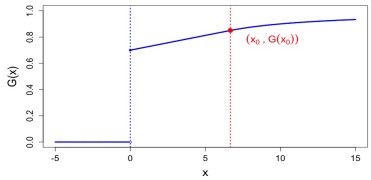


Figure 2: Distribution function G for $p_0 = 0.7$.

An illustration of G for $p_0 = 0.7$ is provided in Figure 2. Next, we choose the distribution function $F_{(s,t)}$ of $X(s,t)$ as the empirical distribution function $F_{(s)}$ (i.e., at each grid cell s) when $X(s,t) \leq u(s,t)$, and according to (6) when $X(s,t) > u(s,t)$. We use spatial models for the marginal tail parameters, whose estimators $\hat{\mu}(s)$, $\hat{\sigma}(s)$ and $\hat{\gamma}(s)$ in $F_{(s)}$ are obtained by composite marginal likelihood inference (Varin et al., 2011) using a threshold $u(s)$ chosen as a high empirical quantile for fixed s; here, we choose the 0.95-quantile of hourly rainfall intensities. Thanks to the consistency of these estimators and the continuity of T, we can apply the continuous mapping theorem such that the transformation $\hat{T}$ (with estimators plugged in) provides a consistent estimate of T.

5.3 Choice of spatio-temporal cost functionals

Our first cost functional $\ell^{(1)}_{s,t}$ is a *spatio-temporal average*, i.e., the average value of $X^*(s,t)$ over the spatio-temporal neighborhoods $\mathcal{N}(s,t)$, for all feasible space-time points (s,t). In space, we specify this neighborhood through a 15 km disc centered at s; in time, it extends backward from t such that $\mathcal{N}(t) = \{t-(\delta-1), \ldots, t\}$ with duration $\delta = 12$ hours. With this choice, $\theta_{\ell^{(1)}_{s,t}} = 1$, see Sections 3.2 and 4.3. The second cost functional $\ell^{(2)}_{s,t}$ is in line with classical EVT and is called *spatio-temporal maximum*; the value $\ell^{(2)}_{s,t}(X^*)$ corresponds to the maximum over the whole study area and observation windows, i.e., $\mathcal{N}(s) = S$, and $\ell^{S}_{s,t} = \max$, $\ell_T = \max$. For this choice, we need an estimate of the extremal coefficient $\theta_{\ell^{(2)}_{s,t}}$ to obtain return levels for lifted events. Therefore, we implement maximum censored likelihood for estimating the scale parameter $\theta_{\ell^{(2)}_{s,t}}$ of a Pareto distribution with fixed shape 1, using observed magnitudes $\ell^{(2)}_{s,t}(X^*)$ censored below a high threshold u.

5.4 Analysis of extremal dependence properties

Using techniques proposed in Section 3.3, we first illustrate pairwise empirical extremal coefficients with respect to spatial distance and temporal lags, and we then check if threshold stability is a valid assumption for the data set when considering high quantiles.

Figure 8 shows the estimations of the empirical spatial and temporal extremal coefficient functions. The pairwise estimator is based on a threshold for each of the two components, see Appendix B for details. The empirical spatial extremal coefficient function is plotted in Figure 8 (left panel). For θ^{spa}, the threshold $u(s)$ is set to the empirical 0.98-quantile of $X^*(s,\cdot)$ where s represents the site with maximum empirical 0.98-quantile between the two sites involved the pairwise estimator. The empirical temporal extremal coefficient function is plotted in Figure 8 (right panel). In this case, a uniform threshold u is chosen as the empirical 0.98-quantiles of the sample of spatio-temporal maxima $\ell^{(2)}_{s,t}(X^*)$. Pointwise

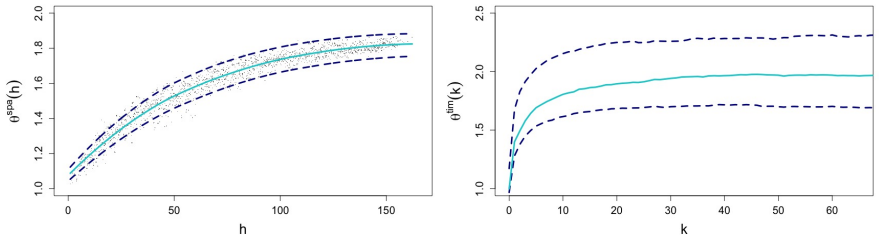


Figure 3: Extremal coefficient functions. Left: $\hat{\theta}^{spa}(h)$, based on a subsample of 1500 pairs of grid cells, with a local polynomial regression (turquoise line). Right: $\hat{\theta}^{tim}(k)$, based on pairs of spatial maxima separated by a time lag k (turquoise line). Dashed lines show bootstrap confidence intervals at 95%.

block bootstrap confidence intervals at 95% for both extremal coefficient function are constructed using variable size blocks with block length following a geometric distribution with mean 300 hours (Politis and Romano, 1994; Davis et al., 2011). Figure 8 shows that $\hat{\theta}^{spa}(h)$ and $\hat{\theta}^{tim}(k)$ always remain below 2 for all spatial distances h and for time lags k lower than 12 hours, hinting at substantial extremal dependence at finite, observed quantile levels. Therefore, we see that the maximum duration of extreme episodes is approximately 12 hours. We point out that there is a certain sensitivity of the estimated curves with respect to the probability p used for fixing empirical thresholds u, with a slight tendency towards decreasing dependence strength at higher levels; see Figure 8 in Appendix B.

Next, we complement these findings by checking spatial threshold stability based on the independence of scales and profiles for high event magnitudes observed at a given time. First, we point out that certain calculations for extreme episodes were quite sensitive to the high proportion of 0 values (i.e., absence of precipitation) in the data set, which amount to around 92 %. Therefore, we add a preprocessing step where we remove hourly time steps t_i from the data set if the precipitation totals in a sliding 24hour-window centered at t_i, cumulated over all grid cells, are smaller than 550 mm, corresponding to a spatially averaged precipitation total of 0.05 mm over 24 hours. The resulting data subset retained contains only around 23 % of 0 values. Now, we check the scale-profile independence for the case of the spatio-temporal average $\ell^{(1)}_{s,t}$, and for simplicity we here consider the full study area $\mathcal{S}$ as spatial support, and we write $\ell^{(1)}_t$ for the resulting cost functional. The empirical 0.95-quantile of $\ell^{(1)}_t(X^*)$ is used as threshold u. Denote by $Y_i = \{Y_i(s,t)\}$ the observed profile process Y_t corresponding to each extracted extreme episode i, with $i = 1, \ldots, m$. In the two displays on the left of Figure 4, the proportion of profile process values $Y_i(s,t)$ below or equal to a threshold $u' \geq 0$, denoted by $f_{u'}(Y_i)$, is plotted for $u' = 0$ and for u' fixed to the empirical 0.95-quantile of all episodes Y_i taken together. The empirical standard deviation $sd(Y_i')$ of the square root $Y_i'(s,t)$ of profile process values $Y_i(s,t)$ is depicted in the third display of Figure 4. For easier visual interpretation, both summary statistics are plotted against $1 - u/\ell^*_i$, where $\ell^*_i = \ell^{(1)}_{t_i}(X^*_{[i]})$. Under the functional domain-of-attraction assumption, the distribution of $1 - u/\ell^*_i$ is approximately uniform on $[0,1]$. A QQ-plot of $1 - u/\ell^*_i$ is shown in the fourth display of Figure 4 with pointwise confidence bounds, and no striking deviation from uniformity appears. Moreover, it is difficult to detect strong systematic trends in profile values with respect to event magnitude. Judging from the shape of the local regression curves in this plot, e.g. for $sd(Y_i')$, this may be a border case between asymptotic dependence and asymptotic independence, but it is difficult to decide with certainty. If data

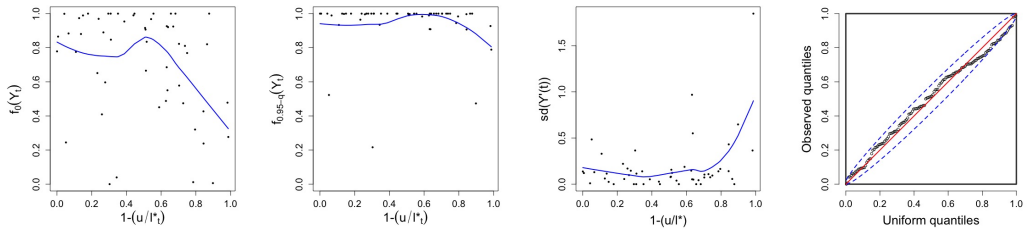


Figure 4: Analysis of scale-profile independence for the spatio-temporal average. The threshold u is chosen as the 0.95-quantile of observed magnitudes $\ell_t^{(1)}(X^*)$. From left to right: $f_{u'}(Y_i)$ for $u' = 0$; same for $u' = 0.95$-quantile; $\mathrm{sd}(Y_i')$; QQ-plot of observed $1 - u/\ell_i^*$ against uniform theoretical quantiles with pointwise confidence interval at 95% (dashed lines).

do not satisfy asymptotic dependence, we acknowledge that our resampling procedure may lead to rather conservative estimates of aggregated extreme risks. In the following, we assume that domain-of-attraction properties are satisfied for our data set if we fix the 0.95-quantile as threshold for $\ell_{s,t}^{(1)}(X^*)$ given as the spatio-temporal average function. Similar conclusions are valid for the spatio-temporal maximum $\ell_{s,t}^{(2)}$ with u given as the 0.98-quantile.

5.5 Parameter choice for extreme episode extraction and lifting

As before with $\ell_{s,t}^{(1)}$, we fix the duration of extreme episodes to 12 hours (i.e., $\delta = 12$ in Algorithm 1). The spatial neighborhoods $\mathcal{S}'$ are chosen differently for the two ℓ functions. In order to calculate $\ell_{s,t}^{(1)}$, we consider a spatial neighborhood of 15 km around each reference point s, such that $\mathcal{S}'$ is composed of sites s with minimum distance of 15 km to the boundary of the study region. However, for $\ell_{s,t}^{(2)}$, we always take $\mathcal{S}' = \mathcal{S}$. We set $\beta = 1$ to separate extreme episodes by at least 1 hour. In order to illustrate a strong uplifting effect in resampled extreme episodes, we select a high lower threshold for newly sampled scale variables R_i, i.e. a large scale parameter α for Pareto distribution with shape 1, here given by twice the maximum value of observed magnitudes $\ell_i^{(1)}(X^*)$ and $\ell_i^{(2)}(X^*)$. In general, the parameter choice in our method provides high flexibility with respect to the modeling context.

5.6 Spatio-temporal extreme precipitation scenarios

We report the ending time t_i for the 6 most extreme precipitation episodes with respect to the spatio-temporal average $\ell_{s,t}^{(1)}$ in the second column of Table 1. Analogously, for the spatio-temporal maximum functional $\ell_{s,t}^{(2)}$ the ending times t_i are presented in the third column Table 1.

In general, we remark that both cost functionals extract similar extreme episodes in terms of temporal neighborhoods, but the order with respect to event magnitudes is different. Some extreme episodes arise only for one of the two cost functionals. In addition, we notice that extreme precipitation scenarios are more frequent during the months of September and October.

Figure 5 shows the original precipitation data $X(s,t)$ and the final uplifted scenarios $W(s,t)$ for

Table 1: Ending times of the most extreme, temporally declustered space-time episodes extracted by considering two different cost functionals.

Episode	**Spatio-temporal average** $\ell_{s,t}^{(1)}$	**Spatio-temporal maximum** $\ell_{s,t}^{(2)}$
1st	2005-09-06 23:00:00	2005-09-07 01:00:00
2nd	1999-09-03 18:00:00	1999-09-14 10:00:00
3rd	2006-10-12 00:00:00	1999-08-28 22:00:00
4th	2002-09-08 22:00:00	1999-09-03 15:00:00
5th	1999-10-18 07:00:00	2001-07-06 05:00:00
6th	2001-07-06 02:00:00	2006-10-12 02:00:00

several time steps from the extracted temporal neighborhoods for the spatio-temporal average $\ell_{s,t}^{(1)}$. We see a clear increase in intensity in the uplifted precipitation fields in Figure 5. Analog plots for $\ell_{s,t}^{(2)}$ are presented in Figure 6.

5.7 Risk analysis

Risk is a complex notion and can take on a variety of forms with diverse applications. The conventional risk measure in hydrology is that of the univariate return level at probability level $q \in [0, 1]$, denoted as Q_q. A *return level* is a quantile, defined as the magnitude of the event that is exceeded with a probability $1 - q$; then, $1/(1 - q)$ is the associated *return period*. However, the return level fails to give any information about the thickness of the tail of the distribution function. In order to prevent the above shortcoming, an alternative risk measure was proposed in actuarial sciences, the so-called Conditional Tail Expectation (CTE) (Denuit et al., 2005). Information about the thickness of the tail of the distribution is included in the CTE, defined for a given level $q \in [0, 1]$ and for a random variable X by $CTE_q(X) = E(X|X > Q_q(X))$. In contrast to the return level, the CTE measure verifies the subadditivity property for continuous risks.

We perform a risk analysis that aims at exploring differences in uplifted extreme episodes that can be imputed to the choice of cost functionals and of the fixed lower threshold (i.e., the Pareto scale parameter) used for sampling new scale variables R_i. We consider the 3 largest episodes extracted for each of the two cost functionals $\ell_{s,t}^{(1)}$ and $\ell_{s,t}^{(2)}$, see Table 1. We uplift these episodes using R_i as the 0.25-, 0.5- and 0.75-quantiles of the Pareto random distribution with shape 1 and scale α_i, $i = 1, \ldots, 4$ with α_1 corresponding to the value of the cost functional for the most extreme episode centered at t_1, and then $\alpha_2 = 2\alpha_1$, $\alpha_3 = 3\alpha_1$ and $\alpha_4 = 4\alpha_1$.

Two univariate risk measures – the quantile for a fixed probability (*return level*), and the Conditional Tail Expectation (CTE) – are computed for the original episode X_i and for each uplifted episode W_i, where we first aggregate values of $X(s, t)$ and $W(s, t)$ respectively for each spatial grid cell by taking its temporal average over the 12 time steps.

Figure 7 presents the calculated spatial return levels and CTE, respectively, at the levels 0.98 and 0.99 according to the two cost functionals, and with the four lower bounds of the support α_i for the Pareto-distributed scaling variable. Along the y-axes, we also report the quantiles for each of the three original episodes. We first study the return level measures. Clearly, Figure 7 (first and second columns) shows the higher α leads to higher risk. Furthermore, for both cost functionals we see that the highest risk is attributed to the episode with highest magnitude, the first extreme episode (see first and second columns

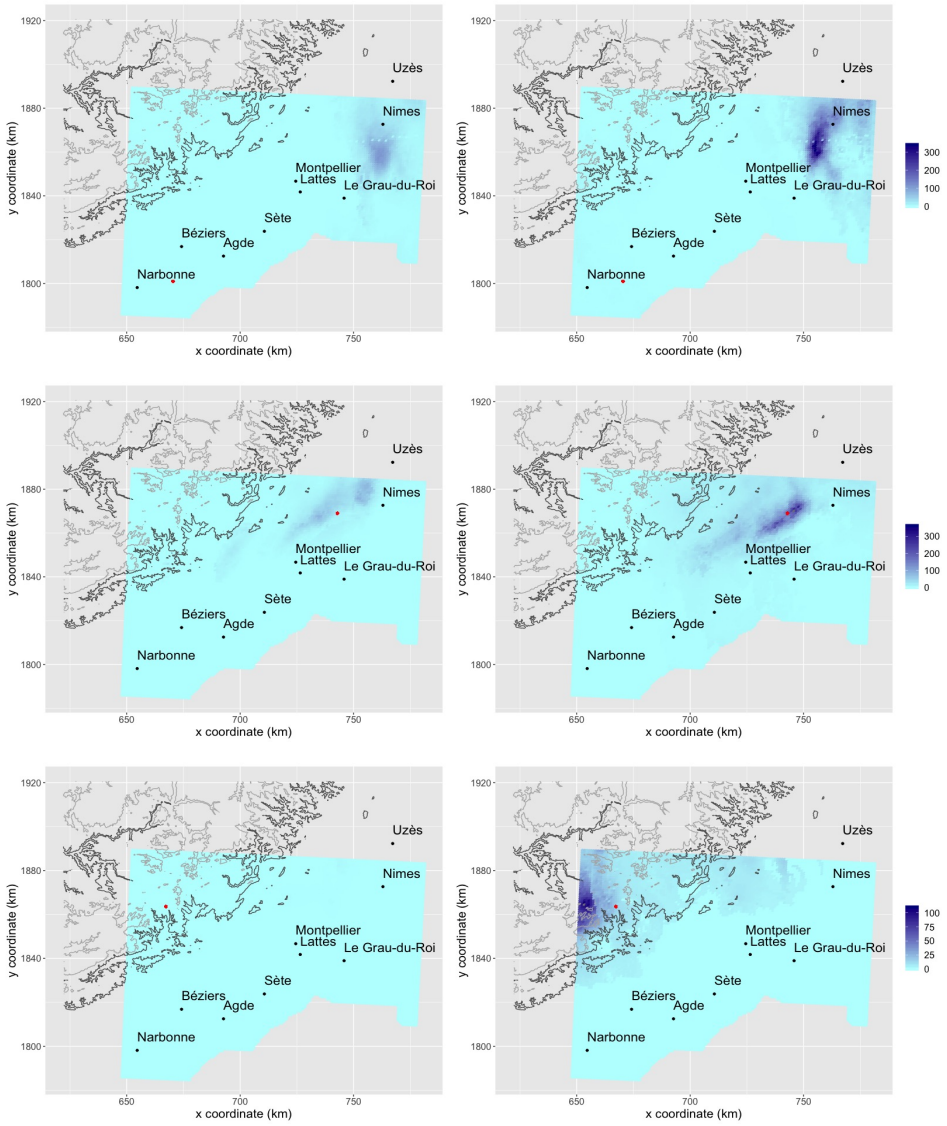


Figure 5: Original precipitation data $X(s,t)$ (left column) and uplifted episodes $W(s,t)$ (right column) based on the spatio-temporal average $\ell^{(1)}_{s,t}$. First row: extreme episode associated to the most extreme episode, here shown for t =2005-09-06, 14:00:00; second row: same for the fourth most extreme episode and t =2002-09-08, 15:00:00; third row: same for the sixth most extreme episode and t =2001-07-06, 00:00:00. The red dots indicate the site s_i where the maximum value $\ell^{(1)}_{s_i,t_i}$ has been observed during the episode. Grey and black contour indicate altitude.

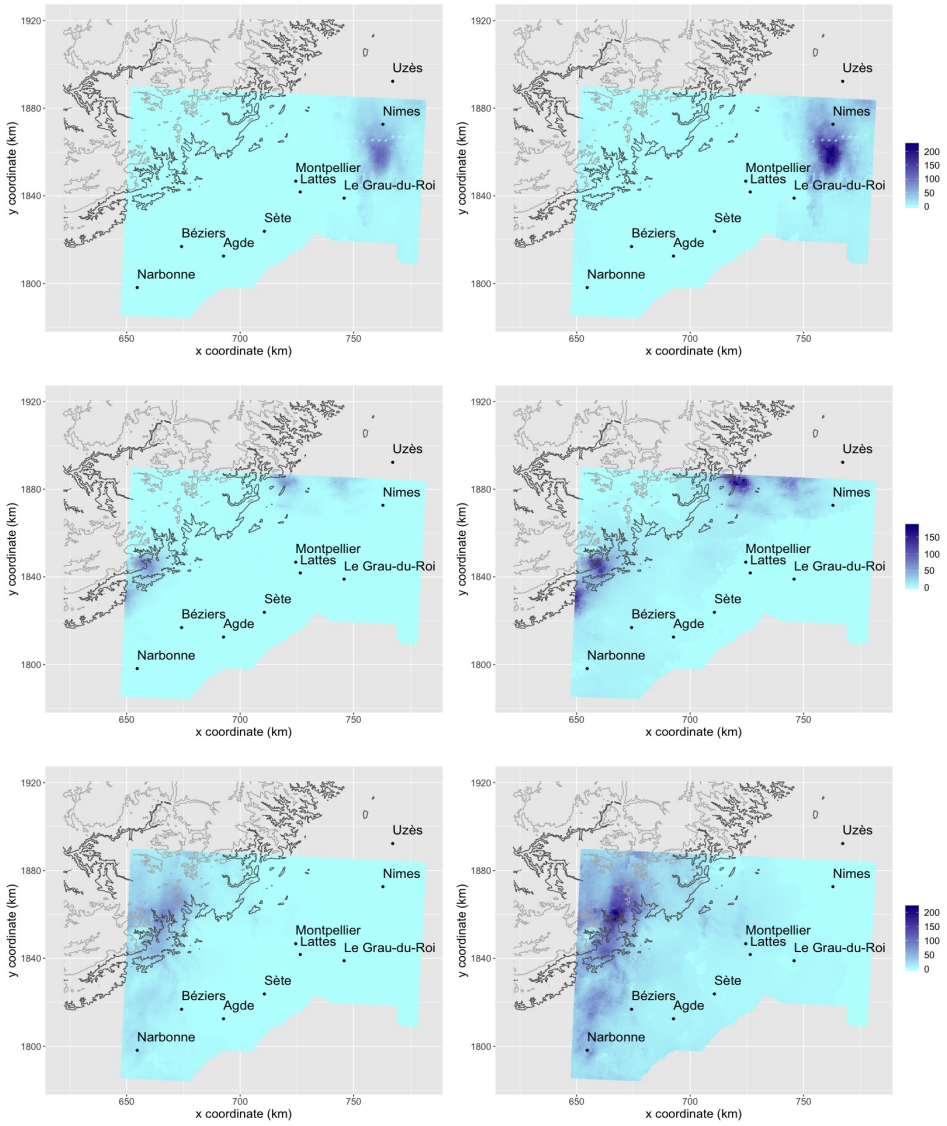


Figure 6: Original precipitation data $X(s,t)$ (left column) and uplifted episodes $W(s,t)$ (right column) based on the spatio-temporal maximum $\ell^{(2)}_{s,t}$. First row: extreme episode associated to the most extreme episode, here shown for t =2005-09-06, 15:00:00; second row: same for the third most extreme episode and t =1999-08-28, 16:00:00; third row: same for the sixth most extreme episode and t =2006-10-11, 20:00:00. Grey and black contour indicate altitude.

in Figure 7). However, the third-highest magnitude event yields higher risk than the second-highest one, as can be seen for the spatio-temporal maximum cost functional (see second row of Figure 7). Indeed, the spatio-temporal maximum ℓ may tend to select episodes with highly localized peaks, i.e. there may be a large majority of zeros or small values with a few spatially confined clusters of very large precipitation intensities. On the other hand, risk measures based on spatio-temporal averages better account for the persistence of moderate to high precipitation intensities. These contrasted results highlight that many ways exist to order elements (here: space-time episodes) defined over high-dimensional spaces (here: space-time neighborhoods $\mathcal{N}(s,t)$); we underline that the mechanism of cost functionals allows the user to make a flexible choice that is appropriate in the modeling context. In addition, in the case of $\ell^{(1)}_{s,t}$, we expect uplifted episodes with the same return periods since $\theta_\ell = 1$ and we use the same realization R (see Section 4.3). Therefore, we obtain greater return levels when the extremeness increases. Similar conclusions are obtained for the CTE risk measure, see third and fourth columns in Figure 7.

6 Conclusion and outlook

In this work, we set up a general framework for space-time generalized Pareto process. It allowed us to develop a semi-parametric method to simulate extreme space-time scenarios of phenomenona such as precipitation. The extremal dependence structure is fully data-driven, and we require parametric assumptions only for the univariate tails, based on EVT. A crucial component is the cost functional defined over a sliding space-time window. It characterizes extreme episodes as episodes whose "cost" exceeds a high threshold. The application of our method to a gridded precipitation data set in Mediterranean France was used for a relatively simple risk analysis. It illustrates how cost functionals can be defined, how these affect the selection of extreme episodes, and how the magnitude of the newly sampled scale variables impacts the magnitude of the lifted extreme episodes on the original marginal scale.

In practice, it is difficult to find extreme value data with long observation periods to empirically study extreme value properties for long return periods without strong modeling assumptions. For practitioners, we provide a methodology that allows them to create extreme scenarios where they can control return levels or periods for aggregated data without any need to explicitly model dependence at extreme quantiles. The proposed methodology requires densely gauged networks or gridded data as spatial interpolation is currently not enabled. Besides precipitation reanalyses, other types of interesting applications include simulations from regional or global climate models.

In future work, space-time distance metrics other than the Euclidean distance could be used to define the space-time neighborhoods $\mathcal{N}(s,t)$. To account for orographic structures, the crossing distance could be used, which includes a vertical component related to the crossing of crests and valleys (Gottardi et al., 2012). Instead of fitting the marginal tail parameters separately for each grid cell, a generalized additive regression approach could be implemented to borrow information from nearby sites (Gardes and Girard, 2010; Carreau et al., 2017). In addition, more sophisticated validation methods for POT stability in large dimensions could be studied. Finally, we note that there are events such as karstic aquifer floods where not only the extreme rainfall but also dry and moderate rainfall periods have to be considered. By extending ideas in Cantet et al. (2011) and in Yiou (2014), we plan to implement our method as part of a spatial precipitation generator that simulates complete rainfall series. Rain-flow models will then be fed by simulated series from a precipitation generator, and we will be able to study the impact of the flood by applying risk measures to the outputs of rain-flow models.

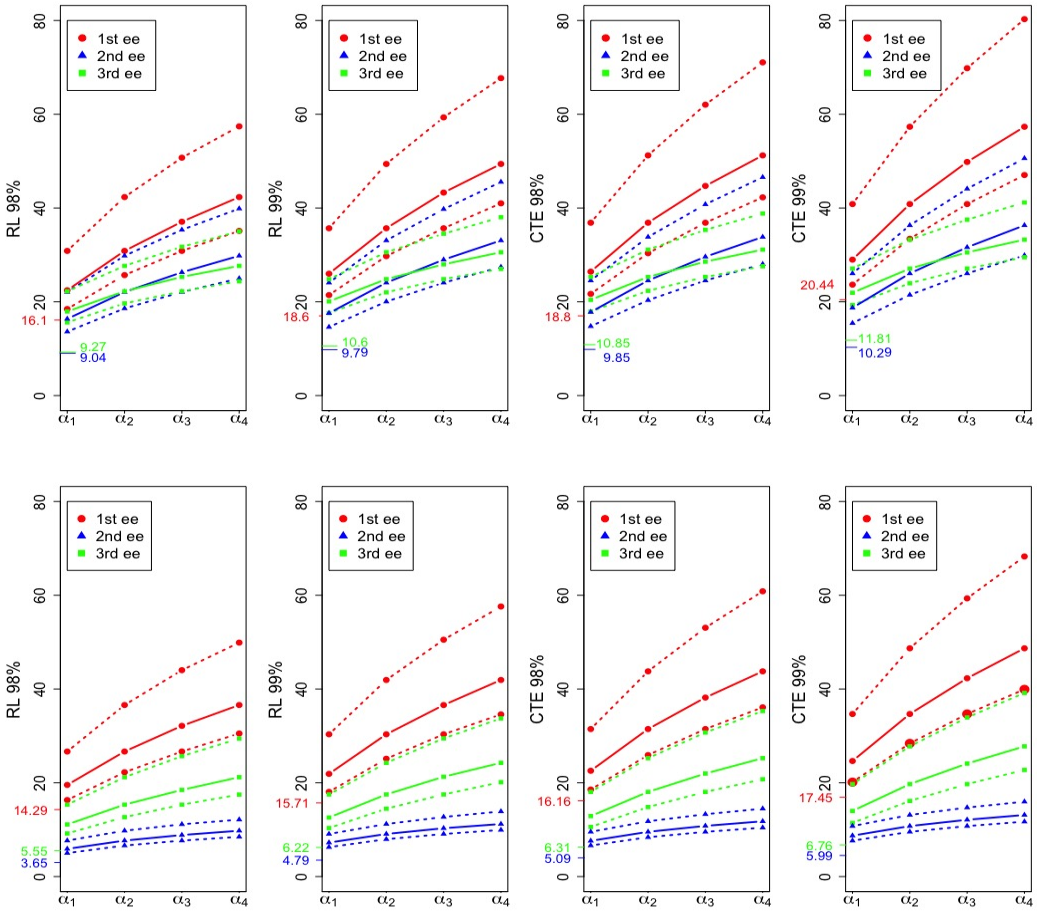


Figure 7: Return level and Conditional-Tail-Expectation at 98% and at 99%. First row: spatio-temporal average cost function. Second row: spatio-temporal maximum cost functional. The legend indicates the extreme episode (ee). For each episode, the lines correspond to different uplifting levels using the 0.25-, 0.5- and 0.75- quantile (from bottom to top) of the Pareto distribution of the scaling variable with shape 1 and scale α_i, $i = 1, \ldots, 4$.

Appendix

A Example: Pareto processes with log-Gaussian profile process

If Gaussian process models are not well adapted to modeling extremes, they can nevertheless be used to construct flexible spatial or spatio-temporal limit models (Kabluchko et al. (2009); Engelke et al. (2015)). For instance, De Fondeville and Davison (2018) analyse the extreme rainfall in the east of Florida by fitting a spatial generalized Pareto process based on log-Gaussian processes. Sample-continuous max-stable process $\{Z(s,t)\}_{s\in\mathcal{S},t\in\mathcal{T}}$ with unit Fréchet margins can be characterized constructively as (de Haan (1984); Schlather (2002))

$$Z(s,t) = \max_{i\geq 1} \xi_i \psi_i(s,t),\ s \in \mathcal{S}, t \in \mathcal{T}, \tag{A.1}$$

where $\{\xi_i,\ i = 1, 2, \ldots\}$ is a point process on $[0,\infty)$ with intensity $\xi^{-2}d\xi$, and $\psi_i(s,t)$ are independent copies of a nonnegative random function with $\mathbb{E}\psi_i(s,t) = 1$, and independent of $\{\xi_i\}$. Specifically, one may choose $\psi_i(s,t) = \exp\{X(s,t) - \sigma^2(s,t)/2\}$ with a centered Gaussian process $\{X(s,t)\}$ possessing variance function $\sigma^2(s,t)$. Regarding the ℓ-Pareto processes equivalent to such max-stable processes, the choice of $\ell(x) = x(s_0,t_0)$ for a fixed space-time point (s_0,t_0) is particularly interesting. In this case, the profile process $Y(s,t)$ in the generalized Pareto process is a log-Gaussian process given by $Y(s,t) \overset{d}{=} \exp\{X(s,t) - X(s_0,t_0) - \frac{1}{2}var(X(s,t) - X(s_0,t_0))\}$ where var denotes the variance. The idea of conditioning on a fixed component of a process is more widely known as the conditional extremes approach (Heffernan and Tawn (2004); Wadsworth and Tawn (2018)), and it arises as a special case of the cost functional ℓ.

B Estimator of extremal coefficient

Let $X^{(1)},\ldots,X^{(M)}$ be identically distributed random variables with unit Fréchet distribution, that is, $\mathbb{P}(X^{(k)} \leq x) = e^{-1/x}$, $x > 0$, $k = 1,\ldots,M$. When the joint distribution of the random vector $(X^{(1)},\ldots,X^{(M)})^T$ follows a multivariate extreme value distribution, then the distribution function of $\max_{k=1}^{M} X^{(k)}$ is $e^{-\theta/x}$, $x > 0$, where $\theta = \theta(X^{(1)},\ldots,X^{(M)})$, $1 \leq \theta \leq M$, is called the extremal coefficient (Smith, 1990; Schlather and de Tawn, 2003). In practice, the coefficient θ can be interpreted as the equivalent number of asymptotically independent random variables (i.e., the effective sample size of extremes) in a random vector $(X^{(1)},\ldots,X^{(M)})$; it quantifies the dependence for extreme values. The case $\theta = 1$ represents full dependence, whereas $\theta = M$ represents full independence.

When considering threshold exceedances, extreme realizations are those that exceed a high threshold. Suppose that $(X_i^{(1)},\ldots,X_i^{(M)})^T$, $i = 1,\ldots,n$ are independent and identically distributed (iid) copies of the random vector $(X^{(1)},\ldots,X^{(M)})^T$, where a threshold exceedance is observed for $X_i^{(k)}$, $1 \leq k \leq M$ if $X_i^{(k)} > u_i^{(k)}$ for some fixed threshold $u_i^{(k)}$; otherwise, the observation $X_i^{(k)}$ is considered as being left-censored at $u_i^{(k)}$. Caires et al. (2011) propose an estimator of the extremal coefficient constructed as

$$\hat{\theta} = m \Bigg/ \sum_{i=1}^{n} \frac{1}{\max(X_i, u_i)} \tag{B.1}$$

where $X_i = \max(X_i^{(1)},\ldots,X_i^{(M)})$, $u_i = \max(u_i^{(1)},\ldots,u_i^{(M)})$, and m is the number of excesses $X_i > u_i$.

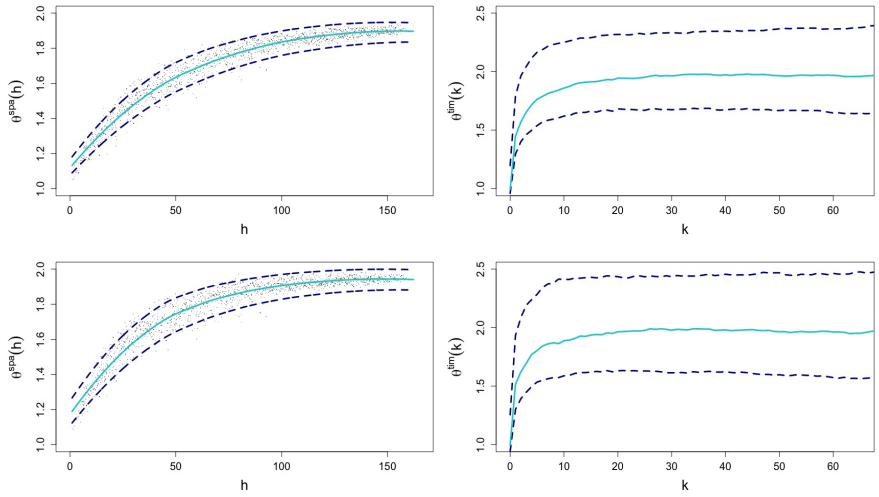


Figure 8: Empirical extremal coefficient functions. Left: $\hat{\theta}^{spa}(h)$, based on a subsample of 1500 pairs of grid cells, with a local polynomial regression (turquoise line). Right: $\hat{\theta}^{tim}(k)$, based on pairs of spatial maxima separated by a time lag k (turquoise line). Bootstrap confidence intervals at 95% (dashed lines). Threshold values u_i in (B.1) are defined as the empirical q-quantile with 0.99 (first row) and 0.995 (second row).

For a summary of extremal dependence with respect to distance in space and time, we follow common practice and focus on pairwise extremal coefficients calculated from bivariate data $X_i = \max(X_i^{(1)}, X_i^{(2)})$ with $M = 2$ in (B.1). We work with two extremal coefficient functions. The spatial extremal coefficient $\theta^{spa}(h)$ measures the extremal dependence between pairs of sites separated by a spatial distance h at a given time. The time extremal coefficient $\theta^{tim}(k)$ measures the dependence between pairs of observations separated by a time lag k at a given site (see Section 2.2 in Chailan et al. (2017) for more details). We estimate empirical spatial extremal coefficient functions from data by considering pairs with structure $(X(s,t_i), X(s+\Delta s,t_i))$ where $\Delta s = h$, while we use $(\max_{s\in\mathcal{S}} X(s,t_i), \max_{s\in\mathcal{S}} X(s,t_i+k))$ for empirical temporal extremal coefficient functions where $\mathcal{S}$ is the study area.

Figure 8 (left panel) presents the spatial extremal coefficient estimates. We set u_i in (B.1) as the maximum of the empirical q-quantiles of $X(s,t_i)$ and $X(s+h,t_i)$, where the latter two variables represent a pair of sites separated by a given spatial distance h at a given hour t_i. The temporal extremal coefficient estimates are plotted in Figure 8 (right panel). In this case, the threshold values u_i are chosen as an empirical q-quantiles of the spatial maximum. The following values for q are used : 0.99 and 0.995 (rows from top to bottom in Figure 8, respectively). Block bootstrap confidence intervals at 95% for both extreme coefficients are constructed by resampling blocks of hours with variable size following a geometric distribution with a mean of 300 hours (i.e. approximatively 12 days) (Politis and Romano, 1994; Davis et al., 2011).

Acknowledgements We thank Météo-France for providing us the data set. We are grateful to the LabEx NUMEV and the French national program LEFE/INSU for financial support.

References

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics.

Brunet, P., Bouvier, C., and Neppel, L. (2018). Retour d'expérience sur les crues des 6 et 7 octobre 2014 à montpellier-grabels (hérault, france) : caractéristiques hydro-météorologiques et contexte historique de l'épisode. *Géographie physique et environnement*, 12:43–59.

Caires, S., de Haan, L., and Smith, R. L. (2011). On the determination of the temporal and spatial evolution of extreme events. *Deltares report 1202120-001-HYE-004 (for Rijkswaterstaat, Centre for Water Management)*.

Cantet, P., Bacro, J., and Arnaud, P. (2011). Using a rainfall stochastic generator to detect trends in extreme rainfall. *Stochastic Environmental Research and Risk Assessment*, 25(3):429–441.

Carreau, J., Naveau, P., and Neppel, L. (2017). Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. *Wat. Resour. Res*, 53(5):4407–4426.

Chailan, R., Toulemonde, G., and Bacro, J. N. (2017). A semiparametric method to simulate bivariate space-time extremes. *Ann. Appl. Statist.*, 11(3):1403–1428.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics.

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013a). Max-stable processes for modeling extremes observed in space and time. *J. Korean Statist. Soc.*, 42(3):399–414.

Davis, R. A., Klüppelberg, C., and Steinkohl, C. (2013b). Statistical inference for max-stable processes in space and time. *J. R. Statist. Soc. B*, 75(5):791–819.

Davis, R. A., Mikosch, T., and Cribben, I. (2011). Estimating Extremal Dependence in Univariate and Multivariate Time Series via the Extremogram. *arxiv:1107.5592v1*.

De Fondeville, R. and Davison, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika*, 105(3):575–592.

de Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Probab*, 12(4):1194–1204.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory. An Introduction.* Springer Series in Operations Research and Financial Engineering. Springer: New York.

Delrieu, G., Nicol, J., Yates, E., Kirstetter, P.-E., Creutin, J.-D., Anquetin, S., Obled, C., Saulnier, G.-M., Ducrocq, V., Gaume, E., Payrastre, O., Andrieu, H., Ayral, P.-A., Bouvier, C., Neppel, L., Livet, M., Lang, M., du Châtelet, J. P., Walpersdorf, A., and Wobrock, W. (2005). The catastrophic flash-flood event of 8-9 september 2002 in the Gard region, France: A first case study for the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory. *Journal of Hydrometeorology*, 6:34–52.

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2005). *Actuarial Theory for Dependence Risks: Measures, Orders and Models*. Wiley.

Dombry, C., Engelke, S., and Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika*, 103(2):303–317.

Dombry, C., Eyi-Minko, F., and Ribatet, M. (2013). Conditional simulation of max-stable processes. *Biometrika*, 100:111–124.

Dombry, C. and Ribatet, M. (2015). Functional regular variations, Pareto processes and peaks over threshold. *Statistics and Its Interface*, 8(1):9–17.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events for insurance and finance.* Springer. Berlin.

Engelke, S., de Fondeville, R., and Oesting, M. (2018). Extremal behaviour of aggregated data with an application to downscaling. *Biometrika: https://doi.org/10.1093/biomet/asy052*.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *J. R. Statist. Soc. B*, 77:239–265.

European Environment Agency (2007). Directive 2007/60/ec of the European parliament and of the council of 23 October 2007 on the assessment and management of flood risks. OJ L. 288: 27-34.

Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737.

Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.*, 43(1):276–298.

Ferreira, A., de Haan, L., and Zhou, C. (2012). Exceedance probability of the integral of a stochastic process. *J. Multiv. Anal.*, 105(1):241–257.

French, J., Kokoszka, P., Stoev, S., and Hall, L. (2018). Quantifying the risk of heat waves using extreme value theory and spatio-temporal functional data. *Computnl Statist. Data Anal.*, 131:176–193.

Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: an application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204.

Gottardi, F., Obled, C., Gailhard, J., and Paquet, E. (2012). Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains. *Journal of Hydrology*, 432-433:154–167.

Guinot, V., Delenne, C., Rousseau, A., and Boutron, O. (2017). Flux closures and source term models for shallow water models with depth-dependent integral porosity. *Advances in Water Resources*, 122:1–26.

Guinot, V. and Soares-Frazão, S. (2006). Flux and source term discretization in two-dimensional shallow water models with porosity on unstructured grids. *International Journal for Numerical Methods in Fluids*, 50.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, 66(3):497–546.

Huser, R. and Wadsworth, J. L. (2018). Modeling spatial processes with unknown extremal dependence class. *J. Am. Statist. Ass., DOI: 10.1080/01621459.2017.1411813*.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab*, 37(5):2042–2065.

Le, P. D., Davison, A. C., Engelke, S., Leonard, M., and Westra, S. (2018). Dependence properties of spatial rainfall extremes and areal reduction factors. *Journal of Hydrology*, 565:711–719.

Lin, T. and de Haan, L. (2001). On convergence toward an extreme value distribution in c[0,1]. *Ann. Probab*, 29(1):467–483.

McPhillips, L. E., Chang, H., Chester, M. V., Depietri, Y., Friedman, E., Grimm, N. B., Kominoski, J. S., McPhearson, T., Méndez-Lázaro, P., Rosi, E. J., and Shafiei Shiva, J. (2018). Defining extreme events: A cross-disciplinary review. *Earth's Future*, 6(3):441–455.

Oesting, M., Bel, L., and Lantuéjoul, C. (2018a). Sampling from a max-stable process conditional on a homogeneous functional with an application for downscaling climate data. *Scand. J. Statist.*, 45(2):382–404.

Oesting, M., Schlather, M., and Zhou, C. (2018b). Exact and fast simulation of max-stable processes on a compact set using the normalized spectral representation. *Bernoulli*, 24(2):1497–1530.

Opitz, T. (2016). Modeling asymptotically independent spatial extremes based on laplace random fields. *Spatial Statistics*, 16:1–18.

Opitz, T., Bacro, J. N., and Ribereau, P. (2015). The spectrogram: A threshold-based inferential tool for extremes of stochastic processes. *Electron. J. Stat.*, 9:842–868.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.*, 3:119–131.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *J. Am. Statist. Ass.*, 89:1303–1313.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized pareto distributions. *Bernouilli*, 12(5):917–930.

Schlather, M. (2002). Models for Stationary Max-Stable Random Fields. *Extremes*, 5(1):33–44.

Schlather, M. and de Tawn, J. A. (2003). A Dependence Measure for Multivariate and Spatial Extreme Values: Properties and Inference. *Biometrika*, 90(1):139–156.

Smith, R. L. (1990). Max-stable processes and spatial extremes. *Preprint. University of Surrey.*

Tabary, P., Dupuy, P., L'Henaff, G., Gueguen, C., Moulin, L., Laurantin, O., Merlier, C., and Soubeyroux, J.-M. (2012). A 10-year (1997–2006) reanalysis of quantitative precipitation estimation over france: methodology and first results. *IAHS Publ*, 351:255–260.

Tawn, J., Shooter, R., Towe, R., and Lamb, R. (2018). Modelling spatial extreme events with environmental applications. *Spatial Statistics, DOI:10.1016/j.spasta.2018.04.007*.

Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika*, 102(4):855–870.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42.

Vinet, F., Boissier, L., and Saint-Martin, C. (2016). Flash flood-related mortality in southern France: first results from a new database. *E3S Web of Conferences 7, article number 06001*, 3:3397–3438.

Wadsworth, J. L. and Tawn, J. A. (2018). Spatial conditional extremes. *https://www.lancaster.ac.uk/ wadswojl/CSE-paper.pdf*.

Yiou, P. (2014). Anawege: a weather generator based on analogues of atmospheric circulation. *Geoscientific Model Development*, 7(2):531–543.